Proteomics as a Tool for the Characterisation of Nosocomial Pathogens

Lakshani K. Rajakaruna

A thesis submitted in partial fulfilment of the requirements of Nottingham Trent University for the degree of Doctor of Philosophy

October 2010

Molecular Identification Services Unit
Department for Bioanalysis and Horizon Technologies
Centre For Infections, Health Protection Agency
London

Copyright statement

Acknowledgments

I would like to thank and convey my deepest gratitude to Prof. Haroun Shah for the immense support, encouragement, guidance and knowledge throughout these four years including my brief sojourn into the world of Caribbean (steelpan) music. Also a big thank you to my academic supervisor, Dr. Graham Ball at NTU, for all his support.

I would like to thank Prof. Saheer Gharbia, Head of DBHT for making it possible for me to undertake the genomic work and Dr. Meeta Desai for her unfailing assistance with the VNTR work and for her patience. My sincere thanks to the team from AnaganosTec, Potsdam, Germany, for their considerate help with MALDI work.

Also, a special thanks goes to Hazel Appleton for carrying out the EM work at HPA, London and Howard Tolley at HPA, Porton, for carrying out the SEM work.

A special thank you to Dr. Dunstan Rajendram, for all his support, encouragement and advice especially with the genomics work and Dr. Vesela Encheva for her support and guidance throughout my PhD. I also wish to extend my gratitude to my colleagues, Gill, Linda, Renata, Ali, Nadia, Caroline, Raju and Julie for all their support in various ways and especially to Min for carrying out the Orbitrap work.

Thank you to all my colleagues at DBHT, especially my fellow PhD students, it was great working with you all over the last few years.

A Big Thank You to Uncle Ajith, without you, this work would not have been possible.

Finally, most importantly, I would like to thank my family; my parents, sisters, in-laws and especially to my sister Gayathri and my brother-in-law Shanka for all their immense support, encouragement, love and patience throughout the years.

This Thesis is dedicated to my loving parents: Amma and Appachchi

Abstract

The aim of the present investigation was to develop and assess the potential of various proteomic approaches for the characterisation of two major nosocomial pathogens, *S. aureus* and *C. difficile* and to further investigate the intraspecies diversity of *C. difficile* using a genotypic approach. The surface-associated proteins of *S. aureus* and intracellular stable ribosomal proteins of *C. difficile* were analysed by MALDI-TOF-MS and the resulting spectra interrogated using two databases *viz.* MMU (Waters®) and SARAMIS™ (AnagnosTec) respectively. A total of 134 clinical *S. aureus* isolates were tested using the MMU database and the MicrobeLynx™ software. All were successfully identified with minor contamination errors that corroborated with 16S rRNA sequencing. By contrast, *C. difficile* isolates were only partially identified (to the genus level) using the MMU database and protocol. Changes in the matrix solution and use of the new database (SARAMIS™) resulted in the correct identification of all *C. difficile* isolates and, detailed ultra-structural studies indicated that intracellular proteins were the new diagnostic biomarkers. The cytosolic/membrane-bound proteins of *S. aureus* and *C. difficile* were investigated using SELDI-TOF-MS and, potential biomarkers for MRSA and MSSA studied using Artificial Neural Networks. Seven key ions were detected for predicting MRSA and MSSA correctly. 1D- gel electrophoresis was also carried out on both taxa. To detect novel loci to differentiate the different ribotypes (027 and 001) and other ribotypes of *C. difficile* a molecular method, 'VNTR', was undertaken using 92 isolates belonging to three ribotype groups. Ten novel loci were detected and could be used to differentiate between isolates of different ribotypes belonging to 027 and 001. This is now being implemented as an epidemiological tool because of its high reproducibility and throughput for genotyping of the invasive ribotype 027 and 001 of *C. difficile*.

Abbreviations

µg	microgramme
µl	microlitre
µm	micrometer
2D	Two-Dimension
3D	Three Dimensional
AAD	Antibiotic associated diarrhoea
CAN	Acetonitrile
AFLP	Amplified Fragment Length Polymorphism
AI	Artificial Intelligence
AIDS	Acquired Immune Deficiency Syndrome
ANN	Artificial Neural Network
AP-PCR	Arbitrarily Primed Polymerase Chain Reaction
AUC	Area under the curve
BP	Back Propagation
BSA	Bovine serum albumin
CA-MRSA	Community Acquired Methicillin-Resistant
CBA	Columbia Blood Agar
CC	Clonal Complex
CDAD	*Clostridium difficile* Associated Disease
CMBT	5-Chloro-2-MercaptoBenzoThiazole
Da	Daltons
DHB	2, 5-Dihydroxy benzoic acid
DIGE	Difference in Gel Electrophoresis
DNA	Deoxyribonucleic Acid
DTT	Dithiothreitol
EDTA	Ethylene diamine tetraacetic acid
ELISA	Enzyme Linked Immunosorbent Assay
EMRSA	Epidemic Methicillin Resistant *Staphylococcus aureus*
ESI-MS	Electro Spray Ionisation Mass Spectrometry
FAA	Fastidious Anaerobic Agar
FAB	Fast Atom Bombardment
GDH	Glutamate deHydrogenase
GPI	Gram positive identification

H	hours
HA-MRSA	Hospital Acquired Methicillin Resistant *Staphylococcus aureus*
HCEC	Human Colonic Epithelial Cells
HPA	Health Protection Agency
HPLC	High Performance Liquid Chromatography
ICM-MS	Intact cell MALDI Mass Spectrometry
IEF	Iso electric Focusing
kDa	kilo Daltons
LD	Laser Desorption
m:z	mass to charge ratio
MALDI-TOF-MS	Matrix Assisted Laser Desorption/Ionisation Time of Flight Mass Spectrometry
MIRUs	Microbial Interspersed Repeat Units
MLEE	Multilocus Enzyme Electrophoresis
MLP	Multi Layer Perception
MLST	Multi locus Sequence Typing
mM	mili molar
MMU	Manchester Metropolitan University
MPRL	Microbial Pathogencity Research Laboratory
MRSA	Methicillin resistant *Staphylococcus aureus*
MS	Mass Spectrometry
MS/MS	Mass Spectrometry/ Mass Spectrometry
MSSA	Methicillin sensitive *Staphylococcus aureus*
NAP	North American Pulse Field Type
NCBI	National Centre for Biotechnology Information
NCTC	National Collection of Type Cultures
PAGE	Poly Acrylamide Gel Electrophoresis
PBS	Penicillin binding protein
PCR	Polymerase Chain Reaction
PCA	Principle component analysis
PD	Plasma Desorption
PE	Processing Element
PFGE	Pulse Field Gel Electrophoresis
PMC	Pseudomembranous colitis
PMSF	Phenyl Methyl Sulphonile Fluoride
Ppm	parts per million

PVL	Panton Valentine Leukocidine
PY-MS	Pyrolysis Mass Spectrometry
rDNA	ribosomal Deoxyribose Nucleic Acid
REA	Restriction Endonuclease Analysis
RFLP	Restriction Fragment Length Polymorphism
RMS	Root Mean Square
RNA	Ribose Nucleic Acid
ROC	Receiver operating curve
rRNA	ribosomal Ribose Nucleic Acid
SARAMIS	Spectral ARchive And Microbial Identifications System
SAX/Q10	Strong Anion Exchange
SCCmec	Staphylococcal Cassette Chromosome
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SEAC	Surface Enhanced Affinity Capture
SELDI-TOF-MS	Surface Enhanced Laser Desorption/Ionisation Time of Flight Mass Spectrometry
SEM	Scanning electron microscopy
SEND	Surface Enhanced Neat Desorption
SEPAR	Surface Enhanced Photoliable Attachment and Release
SSM	Slipped Strand Misparing
SSSS	Staphylococcal Scaled Skin syndrome
SSRs	Short Sequence Repeats
ST	Sequence Type
SVM	Support vector machines
SWP	South West Pacific
TFA	Tri-fluoro acetic acid
TOF	Time-of-Flight
UV	Ultra Violet
VISA	Vancomycin-intermediate *S. aureus*
VNTR	Variable Number Tandem Repeats
VRSA	Vancomycin resistance *S. aureus*
WCX/CM10	Weak Cation Exchange

Contents

Chapter 1

Introduction

1 Introduction

1.1 Microorganisms used in the study

Nosocomial infections are mainly acquired or associated with hospitals and are also referred to as hospital-acquired infections or hospital-associated infections (HAI). The importance of these infections cannot be overstated, as it poses a major threat all over the world resulting in high rates of mortality and morbidity. It is also a major financial burden to the health care system due to the increased cost of patient treatment and infection control (Breathnach 2005). Most commonly, the microorganisms involved in these infections are *Staphylococcus aureus* and *Clostridium difficile* both of which arose from the normal flora of man. *S. aureus* is mainly involved with bacteraemia caused by methicillin-resistant *S. aureus* (MRSA) while *C. difficile* is mainly responsible for causing antibiotic-associated diarrhoea (AAD). Because of their significance in disease, these two pathogens are also referred to as "Superbugs" and is the focus of this study.

1.1.1 The genus *Staphylococcus*

The *Staphylococcus* was among the earliest recognised of the pathogenic bacteria, having been characterised in the early 1880s by Rosenbach. He divided the genus in to two species, *S. aureus* and *S. albus.* Based on cell morphology and type of cell aggregation, the *Staphylococcus* was placed together with the genus *Micrococcus.* However the results of DNA base composition and later on DNA-rRNA hybridisation and comparative oligonucleotide cataloguing of 16S rRNA, have indicated that genus *Staphylococcus* is most closely related to the newly described genus *Macrococcus* (Kloos *et al.,* 1998) and is phylogenetically very distantly related to *Micrococcus* .

The genus *Staphylococcus* currently comprises of nearly 40 species and includes some of the most common nosocomial pathogens (Bergey's Manual 2009). Staphylococci are responsible for a plethora of medical problems including skin and soft-tissue infections, surgical site infections, endocarditis and hospital acquired bacteraemia (Casey *et al.,* 2007). Members are

gram positive, spherical cells ranging from 0.5-1.5 μm in diameter and occur singly, in pairs, short chains and grape-like clusters when viewed through a microscope. On blood agar plates they appear as large, round, golden-yellow colonies, often with β-haemolysis. They are non-motile, non-spore forming, facultative anaerobes that commonly colonise the skin of mammals and are widespread in nature. They are usually catalase and coagulase-positive, while most other *Staphylococcus* species are coagulase-negative. Their cell wall contains teichoic acid and peptidoglycan, with L-lysine as the diamino acid a unique interpeptide bridge that is susceptible to lysis by lysostaphin but relatively resistant to lysis by lysozyme.

1.1.2 Clinical significance/ importance of *S. aureus*

S. aureus is a major component of the microbial flora of primates and is a well known human pathogen, which until recent years has been associated with intermittent infections. It colonises the skin and the anterior nares of individuals and is carried by significant proportion of the population (Gordon and Lowy 2008). Colonisation also allows *S. aureus* to be transmitted among individuals in both health care and community settings. Severe infections caused by *S. aureus* are associated with high rates of mortality and morbidity worldwide with significant increases occurring during the 1990s. *S. aureus* infections are often acute and pyogenic and if untreated, may spread to surrounding tissue or via bacteraemia involving other organs. Some of the serious infections caused by *S. aureus* are bacteraemia, pneumonia, osteomyelitis, acute endocarditis, scaled skin syndrome (SSSS) and is a major cause of many serious hospital and community acquired infections (Shopsin and Kreiswirth 2001).

S. aureus possesses a wide range of virulence factors and can cause infections at many anatomical sites. The presence of *S. aureus* in food can be a potential public health hazard since many isolates of *S. aureus* produce enterotoxin (Kaplan and Tenenbaum 1982). Foods commonly associated with the staphylococcal food poisoning are meat, meat products, salads and bakery products. The most common symptoms are vomiting and diarrhoea. Also, Staphylococcal Scaled Skin Syndrome is caused by the effect of exfoliative toxins produced by *S. aureus* (Ladhani *et al.,* 1999) which affects neonates and young children. Toxic Shock Syndrome (Chesney 1981) a community-acquired disease, is also a result of colonisation with *S. aureus* and is characterised by high fever, hypotension and rash (Dinges *et al.,* 2000).

With increased use of antibiotics, the emergence of antibiotic resistance strains of *S. aureus* developed over the years and has emerged as a major global health problem. Reports of MRSA infections occurring in community settings along with high rates of deaths have heightened the public awareness of MRSA (Corriere and Decker 2008). In United States, MRSA is responsible for approximately 25 % of the nosocomial infections and the Press has labelled MRSA as the "Superbug," which killed more people than AIDS in 2005 (Flynn and Cohen 2008). Also the cost for prevention of *S. aureus* infections has a major impact on both the patients and healthcare systems (Corriere and Decker 2008). Due to these factors controlling MRSA remains a primary focus of most hospital infection control programmes (Das and Lambert 2007).

1.1.3 Virulence factors of *S. aureus*

S. aureus has a wide range of virulence factors, both structural and secreted products which can cause infections at many sites (Figure 1.1); expression of these virulence factors occurs only when required by the bacterium. Surface proteins are generally expressed during logarithmic phase whereas secreted proteins such as toxins are produced during stationary phase (Gordon and Lowy 2008). These virulence factors may have several functions in pathogenesis or multiple virulence factors may increase the aggressiveness of strains.

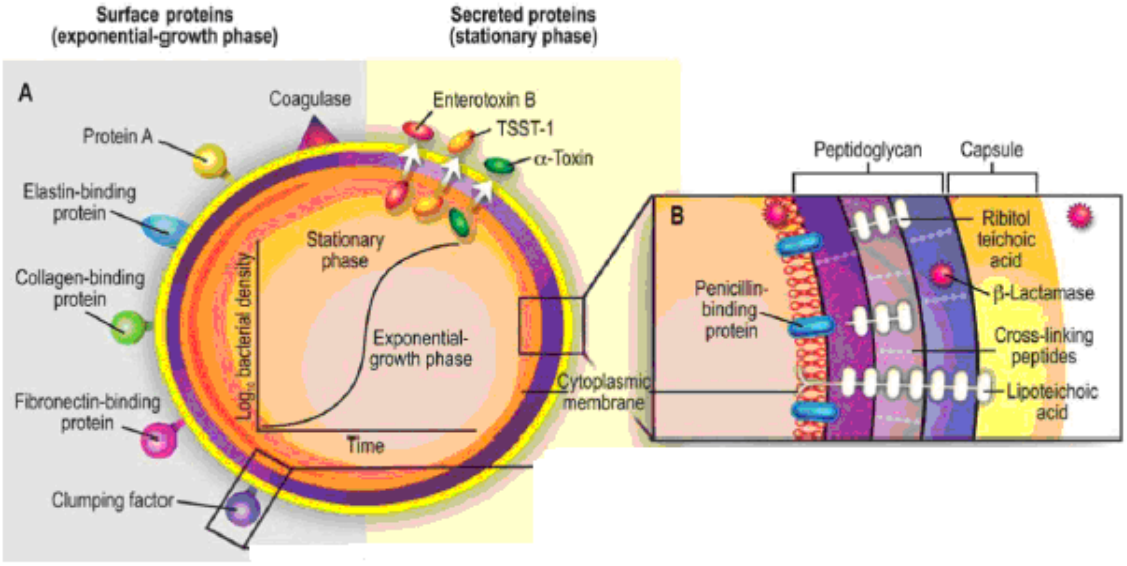


Figure 1.1: Pathogenic factors of *S. aureus* with structural and secreted products both playing roles as virulence factors. A- Surface and secreted proteins. B - Cross section of the cell envelope. Synthesis of many of these proteins is dependent on the growth phase as shown in the above figure (Adapted from Gordon and Lowy 2008).

- ***Surface proteins***

S. aureus has various surface proteins called "microbial surface components recognising adhesive matrix molecules" that mediate adherence to host tissue (Lowy 1998). These proteins bind to molecules such as, collagen and fibrinogen and play an important role in colonisation (Patti *et al.,* 1994). Protein A, the prototype of these proteins has an antiphagocytic property, binds to the Fc portion of the immunoglobulin and results in the disruption of the phagocytosis.

- ***Microcapsule***

Most staphylococci produce a microcapsule which is located externally to the cell wall. Unlike the true capsule of some bacteria, the microcapsule is only visible by electron microscope. *S. aureus* isolates are assigned to 11 capsular serotypes depending on the microcapsular polysaccharides, but the majority of clinical isolates express surface polysaccharides of either serotype 5 or 8 (Lowy 1998). Most MRSA isolates are type 5. Although the function of the capsule in virulence is not known, it's thought to interfere with phagocytosis.

- ***Biofilms (slime)***

Slime is a complex extracellular substance produced in varying amounts by many staphylococci. Once *S. aureus* adheres to the host tissue or prosthetic materials, it can grow and form bioflims which tends to interfere with host defence mechanisms such as opsonisation and phagocytosis (Gordon and Lowy 2008).

- ***Enzymes***

During infection *S. aureus* produces numerous enzymes such as proteases, lipases and elastases that enable it to invade and destroy the host tissues. These bacterial products may facilitate the spread of infection to adjoining tissues (Lowy 1998). Also the enzymes are capable of degrading various macromolecules such as nucleic acid, lipids, proteins and polysaccharides, thereby providing low molecular weight nutrients for growth.

- ***Toxins***

Staphylococci produce numerous toxins that are grouped on the basis of their mechanism of action. *S. aureus* produces a group of pyogenic toxins that cause fever and shock in their hosts. These toxins, also known as superantigens, include enterotoxin and toxic shock syndrome toxin-1 (TSST-1) and are responsible for food poisoning and toxic shock syndrome. Among the *S. aureus* isolates, 20 % are believed to carry the toxin gene *tst* (Lindsay *et al.,* 1998). Another group of toxins called exfoliative toxins are recognised as the cause of SSSS, a disease associated with severe blistering of the skin especially in young children.

1.1.4 Evolution and Epidemiology of *Staphylococcus aureus*

Methicillin-resistant *S. aureus* (MRSA) emerged in 1980s as a major clinical and epidemiologic problem in hospitals. Methicillin, a semi-synthetic penicillin was introduced in the UK in 1961 and was largely developed to overcome the problem of penicillin resistant *S. aureus* in hospitals, which increased throughout the late 1950s. In 1942, two years after the introduction of penicillin for medical use, the first penicillin resistant *S. aureus* was isolated in a hospital (Deurenberg and Stobberingh 2008).The methicillin-resistant *S. aureus* (MRSA), acquired antibiotic resistance as early as 1961, just one year after the launch of methicillin (Jevons 1961). Since then MRSA has gradually disseminated and began causing serious hospital infections worldwide. *S. aureus* is a dynamic and adaptable bacterium that has the ability to acquire antibiotic resistance quickly. MRSA can be resistance to methicillin alone or to one or more antimicrobials used to treat the staphylococcal infections i.e. clindamycine, tetracyclines and gentamycin or to multiple antimicrobial classes (Flynn and Cohen 2008).

One of the most remarkable recent events in chemotherapy was the emergence of vancomycin- intermediate *S. aureus* (VISA) in Japan in 1996 (Hiramatsu *et al.,* 1997). The first vancomycin resistant *S. aureus* (VRSA) strain, containing a VanA resistance operon, was isolated in 2002 in Detroit, MI, the operon having been acquired from a vancomycin resistance enterococcus (Livermore 2000).

S. aureus became methicillin-resistant by acquiring a *mecA* gene, usually carried on a larger segment of DNA called a staphylococcal cassette chromosome *SCCmec* (Katayama *et al.,* 2000; Hiramatsu *et al.,* 2001). In *S. aureus*, the *mec*A gene, coding for the 78-kDa penicillin-binding protein (PBP2a), causes resistance to methicillin and all other β- lactam antibiotics. In methicillin- sensitive *S. aureus* (MSSA), the β- lactam antibiotics bind to the native PBPs that are present in the cell wall, which results in the disruption of the synthesis of the peptidoglycan layer. However in MRSA, because of the presence of the foreign PBP2a, the antibiotics cannot bind and the peptidoglycan layer and synthesis is not disrupted, resulting in the growth of MRSA (Berger-Bachi and Rohrer 2002). The *mec*A gene is regulated by the repressor MecI and the transmembrene β- lactam- sensing signal-transducer MecR1. MecI represses both the transcription of *mec*A and MecR1-MecI in the absence of a β- lactam

antibiotic. In the presence of β- lactam antibiotics, MecR1 becomes active, cleaves MecI, binds to the *mec*A operator region and allows the subsequent production of PBP2a.

The *mec*A gene which is 2.1 kb in length is located on a mobile genomic island called Staphylococcal Cassette Chromosome *mec* (SCC*mec*) (Ito *et al.,* 2003). At present seven main types of SCC*mec* are recognized (Deurenberg and Stobberingh 2008). SCC*mec* type I, IV, VI, and VII causes only β- lactam antibiotic resistance while type II and III cause multiple class resistance due to the presence of additional resistance genes. The origin of SCC*mec* is not known. A study of Wu *et al.* (2001) suggested it was derived from *Staphylococcus sciuri*, a taxonomically primitive Staphylococcal species. In the presence of methicilln, it became resistant due to the increased rate of transcription of the *mec*A homologue (Wu *et al.,* 2001). Another study by Wielders *et al.* (2001) isolated a MRSA strain from a neonate, who had not been in contact with MRSA previously and found the *mec*A gene was identical to that from a *S. epidermidis* strain isolated from the neonate. It was suggested that MRSA had originated *in vivo* through horizontal transfer of the *mec*A gene between the two staphylococcal species (Wielders *et al.,* 2001).

Although MRSA has been seen as a hospital-associated infection, community-acquired MRSA isolates have appeared in recent years (Ho *et al.,* 2004; Nelson *et al.,* 2006). Traditionally, an infection is defined as hospital-acquired (HA-MRSA) if it occurs more than 48 h after admission to the hospital, whereas it is community-acquired (CA-MRSA) if it occurs within 48 h of admission (Bassetti *et al.,* 2009; Flynn and Cohen 2008). HA-MRSA is both genotypically and phenotypically distinct from CA-MRSA. In contrast to HA-MRSA, CA-MRSA isolates are generally susceptible to non ß- lactum antibiotics (Deurenberg *et al.,* 2007; Deurenberg and Stobberingh 2008). HA-MRSA isolates normally contains type I, II or III Staphylococcal Cassette Chromosome *mec* (SCC*mec*) which mediates resistance, where as CA-MRSA isolates contain type IV SCC*mec*. The *S. aureus* population presents a highly clonal structure. The clonality was first discovered by multilocus enzyme electrophoresis (MLEE) and was later supported by multilocus sequence typing (MLST) (Enright *et al.,* 2000; 2002). MLST is a highly discriminatory method of characterising bacterial isolates on the basis of DNA base substitution of several (usually seven) house keeping genes (Maiden *et al.,* 1998).

Historically, HA-MRSA infections have been caused by internationally disseminated clones. A study carried out by Enright *et al.,* 2002 using an international collection of 912 MRSA and MSSA isolates by MLST and SCC*mec* typing, identified 11 major MRSA clones within 5 groups of related genotypes (Iberian, Brazilian, Hungarian, New York/Japan and Pediatric). An MRSA clone was defined as a group of isolates from more than one country which had an identical sequence type (ST) and SCC*mec* type. Enright *et al.* further demonstrated that all major MRSA clones were associated with clonal complexes (CC) CC5, CC8, CC22, CC30 or CC45 and this occurred among isolates with the same MLST type that differ in SCC*mec* type. Also in this study ST8-MSSA, a member of CC8 was observed, as the putative ancestor of the first MRSA strain i.e. ST250-MRSA-I (Figure 1.2). ST8-MSSA is a common cause of epidemic MSSA disease and acquired SCC*mec* type I, II and IV (Figure 1.2). Another clone closely related to ST250 is ST247-MRSA-I; an Iberian clone was one of the major MRSA clones found in European hospitals (Deurenberg *et al.,* 2007). *S. aureus* phage type 80/81, which was responsible for nosocomial and community acquired infections in 1950s is another example of a successful clonal type. ST30 contains a Panton-Valentine Leukocidin (PVL) gene and is related to a southwest Pacific clone, and contains a SCC*mec* IV as well as PVL.

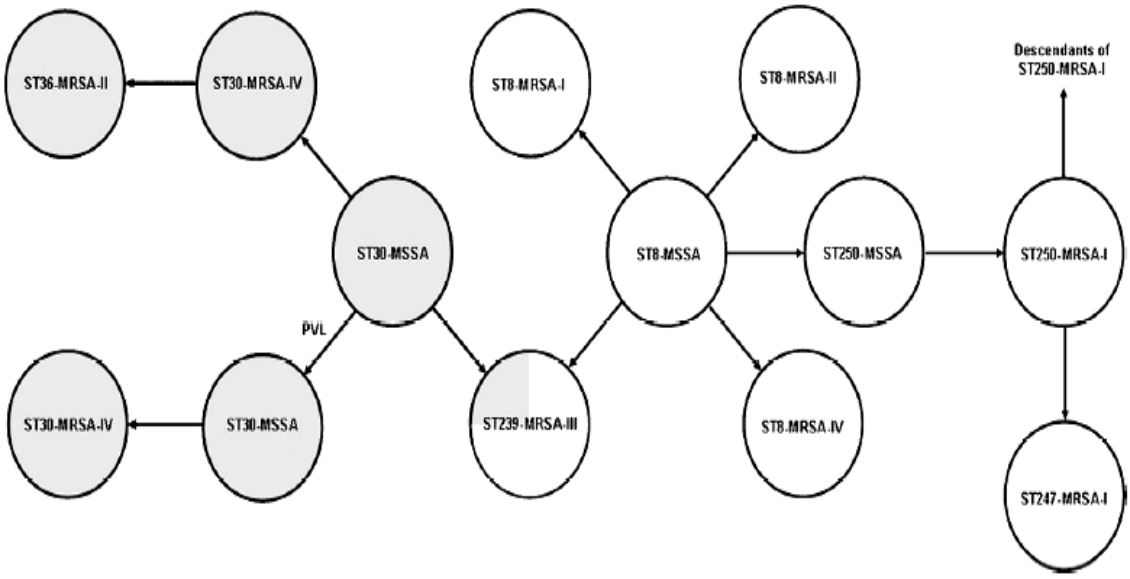


Figure 1.2: The evolutionary origins of the major MRSA clones and the possible relation between CA-MRSA and HA-MRSA. The arrows indicate either (1) the acquisition of SCCmec, (2) a change of SCCmec, (3) a change of ST, or (4) the acquisition of PVL. The grey coloured circles represent MRSA clones from CC30, while the white circles represent MRSA clones from CC8. ST239-MRSA-III from CC8 evolved by the transfer of a 557-kb fragment from the chromosome of ST30 into a ST8 background (Adapted from Deurenburg and Stobberingh 2008).

After the appearance of MRSA in 1960s, in the early 1980s gentamycin was used as a better infection control against for MRSA (Livermore 2000). But soon gentamycin resistance began to emerge in *S. aureus* and MRSA isolates and became a problem in mid-1980s and a strain designated as epidemic MRSA 1 (EMRSA-1) spread widely in the UK. This was followed by another epidemic isolate EMRSA -3, which became prominent in late 1980s and by the 1990s EMRSA 15 and 16. EMRSA-15 and 16 (EMRSA-16) are the most important and prevalent EMRSA isolates found in the UK and have also been found in a number of European countries and the USA. EMRSA-16 is the best described epidemiologically and originated in Kettering, England in 1992 (Murchan *et al.,* 2004). The full genome sequence of this strain has been published (Holden *et al.,* 2004). These two isolates have achieved a major spread and are associated with severe infections (Gould 2008). Apart from the resistance to β lactams, EMRSA-15 is resistant to erythromycin and often resistant to ciprofloxacin. EMRSA-16 is usually resistance to erythromycin, ciprofloxacin and occasionally to gentamicin (Livermore 2000).

CA-MRSA also emerged as a major threat to the community since its prevalence was predicted to rise to 25 % (Chambers 2001). A major outbreak of CA- MRSA was first reported in 1980s in Detroit among the intravenous drug users. Since then, CA-MRSA has gradually increased even without a history of intravenous drug use (Corriere and Decker 2008) where infections frequently occur in young and healthy individuals. With four pediatric deaths from CA-MRSA in Minnesota and North Dakota in 1997-1999, the potential of CA-MRSA become more evident (Centres for Disease Control and Prevention: 1999) even without any risk factors. The strain responsible for these deaths was later termed as MW2 was a ST1 and PFGE type USA400 (Gordon and Lowy 2008 and Said-Salim *et al.,* 2003). DNA sequencing of the this strain later revealed a unique SCC*mec* type IV in contrast to SCC*mec* type I, II and III found in nosocomial MRSA (Said-Salim *et al.,* 2003). Subsequently, outbreaks of skin and soft-tissue infection caused by CA-MRSA were also reported among prison inmates, soldiers and athletes (Tenover and Goering 2009). The strain responsible for these infections was identified as ST8 and PFGE type USA300 (McDougal *et al.,* 2003; Tenover and Goering 2009). Currently this strain was identified as the major cause of hospital infections in USA (Gould 2008). Also a number of other CA-MRSA clones were found to be prevalent. These included ST80 (France-Switzerland), ST30 (SWP clone) and ST93 (Australia Queensland clone) (Vandenesch *et al.,* 2003). A recent outbreak of CA-MRSA appears to be caused by isolates that also carry genes for Panton-Valetine leukocidin (PVL), a toxin that is known to cause lysis of white blood cells. It has been shown that PVL contributes to severe haemolytic and necrotic pneumonia in children (Lina *et al.,* 1999).

1.1.5 Identification of *Staphylococcus aureus* and other staphylococci

With the increasing outbreaks of Staphylococcal and MRSA infections all over the world, controlling these infections remains a primary focus of most hospital infection control programs. As a result bacterial strain typing or subspeciation has become an important clinical tool to investigate suspected outbreaks and to evaluate nosocomial transmission. To date numerous techniques have been implemented to differentiate *Staphylococcus* species and specially MRSA. On the basis of a variety of conventional phenotypic characters, such as colony size, colony pigment, aerobic or anaerobic growth, clumping factor and oxidase production etc. most of the *Staphylococcus* species can be identified (Kloos 1990). However these characters such as biochemical reactions, physiological or morphological reactions, and analysis by conventional methods often require three to five days. To overcome these tedious procedures in routine laboratories, several manufactures have developed rapid identification kits and automated systems requiring only a few hours to one day for completion of tests which provides both rapid and accurate results.

Some of the kits include the API Staph-Ident, DMS Staph-Trac kits, MicroScan Pos ID panel and the fully automated VITEK system which utilize a gram positive identification card (GPI) (Kloos and George 1991). These kits consist of trays or strips with wells containing dehydrated substrates, biochemicals or nutrient media, the wells providing a reaction vessel for enzyme activity. After inoculation, the trays are incubated for five to 24 h and then interpreted. The API Staph-Ident kit (Analytab Products, Plainview, N.Y.) was developed for the identification of coagulase-negative-staphylococci. The main advantage of this kit is it only requires a five hour incubation period (Overman and Overley 1990). A study by Kloos and Wolfshohl (1982) using API Staph-Ident kit resulted in more than 90 % similarity between conventional methods and the kit while a study by Overman and Overley (1990) resulted in 75 - 100 % reproducibility between the two versions of the kits.

Another commercial kit, the DMS Staph-Trac kit (DMS Laboratories, Inc., Flemington, N.J.), a miniaturized biochemical test system, requires a 24 h incubation period. Using this system, Giger and co workers (1984) have demonstrated that up to an 88 % correct identification could be achieved using this system. The MicroScan Pos ID and Rapid Pos ID panel systems (Baxter Diagnostic Inc., MicroScan Division, West Sacramento, Calif.) uses 18 and 42 tests which

could be read after 15-48 h and two hr of incubation respectively where results could be read using a auto Scan WalkAway. Studies by Kloos and George (1991) using this system demonstrated that high degree of similarity could be achieved compared to the conventional systems while Iorio *et al.,* 2007 demonstrate a lower percentage (79.3 %) of correct identification compared to the conventional methods.

The VITEK system is used for identification and susceptibility testing of both gram negative and gram-positive organisms. It contains 25 conventional biochemicals, three proprietary substrates, and one antibiotic and identification cards which are read on the automated reader incubator. The final identification is usually available after 4 to 18 h of incubation (Rhoads *et al.,* 1995). For the identification of *S. aureus*, tests such as the rapid commercial slide test for detecting haemagglutination by clumping factor or the latex agglutination tests to detect clumping factor or protein A could be used. With these systems, identification of most staphylococcal spp. can be made with an accuracy of 70-95 % (Kloos 1990).

To ascertain the correct antimicrobial therapy for treatment of staphylococcal infections, it is necessary that methicillin resistance is detected as rapidly as possible. Many different phenotyping and genotyping methods have been developed to distinguish *S. aureus* isolates (Sousa and Lencastre 2004). The oxacillin agar screening test has been a mainstay for the detection of MRSA in diagnostic laboratories over many years which included oxacillin-resistant screening agar base with 5.5 % NaCl and 2 mg/L oxacillin (Casey *et al.,* 2007). Other techniques such as phage typing, MLEE (Weller 2000), pulsed-field gel electrophoresis (PFGE) (Murchan *et al.,* 2003), amplified fragment length polymorphism (AFLP) (Melles *et al.,* 2004), MLST (Enright *et al.,* 2000), *spa* typing and SCC*mec* typing for detection of the presence of *mec*A gene using PCR (Swenson *et al.,* 2001) are carried by different laboratories. All of these techniques enable the sub typing of unrelated isolates with different discriminatory power, accuracy and reproducibility (Melles *et al.,* 2007).

PFGE is one of the most discriminatory and frequently used techniques which became the gold standard for outbreak investigation of *S. aureus* (Weller 2000). In PFGE the *S. aureus* chromosomal DNA is digested with restriction enzyme *Sma*I and the resulting fragments are separated in an agarose gel (Tenover *et al.,* 1995). Despite the efforts to harmonize the protocol and the nomenclature, comparison of interlaboratory results remains difficult (Melles

et al., 2007; Deurenberg and Stobberingh 2008). MLST as described above (section 1.1.4) is based on stable seven housekeeping genes and have proven to be an excellent method to study the evolution of *S. aureus* (Enright *et al.,* 2000). However this method is costly, laborious and time consuming. AFLP on the other hand is a whole genome typing method that detects genetic variation between isolates using rear-cutting restriction endonucleases followed by separation of fragments on an agarose gel is more amenable (Melles *et al.,* 2007). Typing of isolates using MLEE involves the extraction of enzymes from the bacterial cell, separation by electrophoresis and examination by selective staining (Weller 2000). This method has also been applied to MRSA typing by Tenover and co workers and achieved good reproducible results (Tenover *et al.,* 1994).

As DNA sequencing and PCR based methods became more assessable and showed higher resolution. The last decade has witnessed wider applications of these methods (Weller 2000). The method developed by Frenay *et al.* 1996 determines the sequence variation of the polymorphic region X of the protein A locus. It's been shown that spa typing could be used for both molecular evolution as well as hospital outbreaks of MRSA (Koreen *et al.,* 2004). Also the detection of the structure of SCC*mec* using PCR has become the most popular method and has developed over the years (Deurenberg and Stobberingh 2008).

1.1.6 Staphylococci infection control

Due to the high rates of mortality and morbidity associated with MRSA worldwide, in both hospital and community, prevention control strategies are becoming more important and receiving increased attention. Multiple infection control practices such as eradication, hand hygiene, identification and isolation of MRSA carriers, patient decolonisation and environmental decontamination are used in combination to prevent nosocomial infections (Gleeson 2008). Also antistaphylococcal immunisation is an area of ongoing research.

There is considerable evidence of *S. aureus* nasal carriage and the link to the staphylococcal infection (Gordon and Lowy 2008). When infections reoccur, studies have shown that patients are generally infected with the same strain from the original infection (Huang *et al.,* 2008). The most commonly used approach to eradicate or decolonise, is the use of topical antibiotics such as mupirocin ointment to the anterior nares twice daily along with chlorohexidine baths, using either 4 % chlorohexidine gluconate containing soap or 2 % chlorohexidine gluconate wipes (Gleeson 2008). Mupirocin, a topical antistaphylococcal agent that inhibits RNA and protein synthesis and eliminates the nasal colonisation in carriers (Wenzel and Perl 1995). Hand hygiene is likely to be the most important component of prevention and control of MRSA in hospitals since HA-MRSA are most commonly transmitted through healthcare workers. Also, isolation of MRSA carriers prevents transmission within the hospital. Another component of controlling HA-MRSA is environmental decontamination. Reports indicate that, 53 % of acquisition of MRSA is derived from the hand of individuals following environmental contact (Bhalla *et al.,* 2004).

At present, no vaccine is generally available that stimulates active immunity against the staphylococcal infections in humans. The most significant antistaphylococal vaccine published to date comprises *S. aureus* type 5 and 8 capsular polysaccharides conjugated to nontoxic recombinant *Pseudomonas aeruginosa* exotoxin A (Robbins *et al.,* 2004). Evaluation by randomised trials on haemodialysis patients showed partial immunity against *S. aureus* bacteraemia for approximately 40 weeks, after which the antibody levels decreased. However a successful vaccine with long term efficacy is still a distant goal and other methods are likely to be used for the foreseeable future.

1.2 ***Clostridium difficile***

The genus *Clostridium* is a large and diverse group of prokaryotes. Members of the genus *Clostridium*, including its pathogenic species are mainly found in soil while some are found as a very small component of the bowel flora of humans and animals. To date more than 190 *Clostridium* species have been described (Bergey's Manual 2009). Because of its diverse and heterogeneous nature, *Clostridium* has become one of the largest bacterial genera clearly in need of taxonomic refinement. The genus *Clostridium* was first proposed by Winslow *et al.* (1920) and *C. butyricum* was proposed as the type species by Cato *et al.* 1986. The genus Clostridia has a wide range of G+C content of 22-55 while the toxigenic species have a narrow G+C content of 24-29 (Hatheway 1990). Based on 16S rRNA gene sequences Collins *et al.,* 1994 has suggested a major revision of the genus. Hence the Clostridia were divided into 19 different clusters where medically important species were included in cluster I.

C. difficile is an anaerobic, gram positive and spore-forming bacterium and microscopically appears as long drumsticks with a bulge located at their terminal ends. It has been isolated from diverse natural habitats, including soil, hay, sand, dung from various large mammals and the faeces of dogs, cats and humans (Lyerly *et al.,* 1988) and is carried asymptomatically as a part of the gastrointestinal flora in many of the healthy new born and in elderly people. *C. difficile* was first isolated by Hall and O'Toole in 1935, who found it as part of the normal microbial flora of the stools of infants (see Louis *et al.,* 1962). They referred to this organism as *Bacillus difficilis* because of the difficulty they encountered in the isolation of the organism. These investigators were also the first to show that the organism is toxigenic.

1.2.1 *C. difficile* infection, epidemiology and risk factors

C. difficile causes a spectrum of nosocomial diseases ranging from Antibiotic-Associated Diarrhoea (AAD) to Pseudo Membranous Colitis (PMC), especially in hospitalized patients those over 65 years (*CDR.,* 2000) and is classically associated treatment with clindamycin, but it may occur after exposure to a wide range of antibiotics such as ampicillin and cephalosporin's because these antibiotics are widely prescribed (Starr 2005; Deneve *et al.,* 2009). There are number of cases reported suggesting that almost any antibiotic can cause the disease (Bouza *et al.,* 2005), and very rarely, the disease can even occur without prior exposure to antibiotics.

C. difficile produces two lethal toxins, enterotoxin and cytotoxin (toxin A and B) and are thought to be responsible for the diarrhoea and inflammation (Poxton *et al.,* 2001). The normal bacterial flora in the gut serves as the major barrier against the colonisation by the pathogen. When the flora is disturbed in some manner, the host becomes susceptible to colonisation or over growth by this pathogen (Cloud and Kelly 2007). Antibiotics such as clindamycin, which is the primary cause of PMC act in this manner by disturbing the normal gut flora of the bowel. Typical symptoms of AAD in older patients are represented by frequent loose watery or bloody stools which maybe accompanied by abdominal cramps and fever (Borriello 1998).

The hospital environment and hospitalised patients are the major reservoirs of *C. difficile* where nosocomial transmission occurs by the faecal-oral route (Doshi *et al.,* 2009). At present epidemic *C. difficile* isolates are widely distributed in the hospital environment, both as a cause and result of nosocomial diarrhoea and it is clear that some isolates are more virulent than others (Wilcox 2003). *C. difficile* is now recognised as the primary cause of hospital-acquired colitis in patients who receive antibiotics, chemotherapeutics or other drugs that alter the normal flora and is also responsible for considerable patient morbidity and cost to the health care system (Wilcox *et al.,* 1996, Dawson *et al.,* 2009). An HPA, Press release (2007) stated that, more than 15,000 cases of *C. difficile* infections occurred in over 65 years during the first quarter of the year. However in the first quarter of 2009 it has decreased to less than 7000 cases (http://www.hpa.org.uk/HPA/). Analysis carried out on discharge patients in hospitals in the US revealed that, CDAD rates increased rapidly from 2001 and doubled until 2003 (McDonald *et al.,* 2006) and is the major cause of health care associated diarrhoea in

USA (Doshi *et al.,* 2009). Due to the increase in *C. difficile* infections, it was reported that in USA, more than $ 1.1 billion costs per annum was spent on management (Kyne *et al.,* 2002). However, *C. difficile* is responsible for only 20 % or less of the AAD while the aetiology of the remaining 80 % is still unknown, although potential suggestions such as *S. aureus* and *C. perfingens* have been proposed (Wilkins and Lyerly 2003 and Bouza *et al.,* 2005).

When the number of CDAD outbreaks began to rise, resulting in increasing number of deaths, the emergence of a hyper virulent epidemic strain was suggested. The earliest report was in Pittsburgh, Pennsylvania (Muto *et al.,* 2005), soon followed by Quebec and Montreal surrounding hospitals in Canada. Later it was found that an uncommon strain of *C. difficile* was responsible for these multiple outbreaks (Loo *et al.,* 2005). It was also been identified as causing outbreaks in UK and in parts of Europe, suggesting, it was a global epidemic strain (Kuijper *et al.,* 2006, Dawson *et al.,* 2009) and it was characterised as toxinotype III, North American Pulsed Field Type 1 (NAP1), restriction enzyme analysis type "B1" and PCR ribotype 027 (Warny *et al.,* 2005). It was found that, in addition to the large clostridial toxins A and B, which are the main virulence factors, NAP1/BI/027 possesses an extra toxin known as binary toxin (Warny *et al.,* 2005). It was also shown that this strain can produce 16 times more toxin A and 23 times more toxin B (McDonald *et al.,* 2005) than any other *C. difficile* strain. In addition, this virulent strain possessed an 18 base pair deletion in gene *tcd*C, a putative regulatory factor of toxin A and B, suggesting that this might be the cause of increased toxin production (McDonald *et al.,* 2005). Whether it was a binary toxin or increase toxin A and B production or other unidentified virulence factors that were responsible for this hyper virulent strain, evidence indicated that this toxinotype III strain was more virulent and caused a major shift in epidemiology of *C. difficile* infection. Apart from the changes in the virulence factors, NAP1/BI/027 also exhibited resistance to fluoroquinolones. An infection with this strain was specifically linked to fluoroquinolone exposure as this was increasingly prescribed (Loo *et al.,* 2005).

It is known that antimicrobial therapy is the most widely recognised risk factor involved in CDAD, by disturbing the normal gut flora and produces a niche for *C. difficile* to multiply. In addition to this, increase use of antimicrobial therapy and use of broad spectrum antimicrobials also contribute to the occurrence of CDAD (Deneve *et al.,* 2009). Other notable

risk factors involved in the infection are the use of stomach suppression medications such as proton pump inhibitors, nasogastric tubes and gastrointestinal surgery (Rupnik *et al.,* 2009).

1.2.2 Virulence factors of *C. difficile*

Toxin A and B

The best described virulence factors of *C. difficile* are the toxins A and B, encoded by the genes *tcd*A and *tcd*B (Warny *et al.,* 2005, Rupnik *et al.,* 2009). Toxin A is an enterotoxin and toxin B is cytotoxin. When the toxins are released in to the bowel, they adhere to the receptors of the colonic brush border and cause necrosis and shedding of these cells into the lumen leaving a shallow ulcer on the mucosal surface (Doshi *et al.,* 2009).

The genes encoding toxin A and B are part of the pathogencity locus (*paLoc*), harbouring five genes (*tcdABCDE*) (Sebaihia *et al.,* 2006) which is a short chromosomal segment carried by pathogenic isolates of *C. difficile* (McDonald *et al.,* 2005). In addition to this, a separate binary toxin has been described in the hyper virulent strain, stated above. The binary toxin is expressed by a minority of *C. difficile* isolates and is encoded in a separate part of the genome (Cloud and Kelly 2007). In the epidemic strain NAP1/BI/027, the role of the binary toxin is still unknown and since its emergence, the prevalence of isolates producing binary toxin has increased (Cloud and Kelly 2007).

1.2.3 Other virulence factors

In addition to the toxins, other factors produced by *C. difficile* may contribute to the colonisation of the gut and its pathology. Adhesion to host cells is known to be important in the expression of virulence for many pathogens.

Studies by Borriello (1990) and Borriello and Wilcox (1998) using various strains of *C. difficile* in a hamster model demonstrated the presence of fimbriae at the pole of the cell. However their role in colonisation remains unclear (Borriello and Wilcox 1998). Additionally, the heat shock protein GroEL in *C. difficile* has been shown to have a role in cell adherence (Hennequin *et al.,* 2001). For many bacteria, the possession of a capsule is an important virulence factor and there is evidence that some isolates of *C. difficile* possess capsule-like material *in vitro* (Borriello 1990). The presence of an antiphagocytic factor in *C. difficile* was reported by Dailey *et al.* (1987) but the removal of the cell surface carbohydrates showed no marked effect on the rate of phagocytosis. The role of hydrolytic enzymes in *C. difficile* infections remains uncertain but it has been shown by Seddon *et al.* (1990) that virulent strains possess a higher level of hyaluronidases.

1.2.4 Diagnosis and identification

Diagnosis of *C. difficile* is mainly done by detection of toxin in stools while the most definitive way of diagnosing PMC is by endoscopic detection of pseudomembranes or micro abscesses in antibiotic treated patients with diarrhoea (Lyerly *et al.,* 1988). Once *C. difficile* has been identified, efforts to minimise its transmission is required to reduce additional cases or outbreaks by isolating the infected patient, followed by disinfection with chemical agents in order to minimise further risks (Wilkins and Lyerly 2003).

As an *in vitro* diagnostic tool for *C. difficile* detection, many clinical laboratories use cycloserine-cefoxitin-fructose agar as a selective medium (George *et al.,* 1979). Isolates are presumptively identified by their colony and cellular morphologies, fluorescence, volatile fatty acid profiles and other properties. In general, many clinical laboratories use antibody-based tests, because these tests are more cost-effective and have reduced turnaround times (*ca.* 15 - 45 min). These tests target the toxin A, a combination of toxin A and B or the common antigen; glutamate dehydrogenase (GDH), which is an essential enzyme, produced by all *C. difficile* isolates and could be readily used to detect the presence of this bacterium in faecal samples. However, antiserum against *C. difficile* cross reacts with other anaerobes and newer GDH tests avoided this potential pitfall by using antibodies only react with *C. difficile* (Wilkins and Lyerly 2003).

Enzyme linked immunosorbent assay (ELISA) which detects the toxin A and a combination of toxin A and B is also used as a diagnostic tool in hospitals (Wilkins and Lyerly 2003). These tests are easier, more rapid, less expensive and exhibit a higher sensitivity (> 80 %), although it's less sensitive than the tissue culture assay. ELISAs that detect toxin A and B were considered unnecessary, until the recent discovery of a *C. difficile* isolate CCUG 8864 (Torres 1991) which is $A^- B^+$, and toxin B was found to be more enterotoxic than toxin B from typical $A^+ B^+$ isolates.

The tissue culture assay has been used extensively for the detection of *C. difficile* toxin in stools specimens. Despite being more time consuming and tedious, the test detects picogram levels of *C. difficile* toxin, making it the most sensitive test available at present and also referred to as the "gold standard" (Wilkins and Lyerly 2003). There are possible problems,

however, with the assay as it is used currently. The assay is not highly standardised, which results in variations in the most appropriate cell line to use and the optimal faecal dilution *etc.*, making it difficult to compare the clinical studies (Lyerly *et al.*, 1988).

1.2.5 Typing of *C. difficile*

Although isolation of toxigenic *C. difficile* from samples is adequate for diagnosis, the typing of isolates is essential for epidemiological outbreak investigations. To understand the nosocomial epidemiology of *C. difficile* infection, various typing or fingerprinting methods have been applied.

Early methods were mainly based on phenotypic properties such as antibiograms, which was studied by Burdon (1982). Although he was able to identify common resistance patterns to three antibiotics, the method was rudimentary and soon was suppressed by other methods. Several other techniques were applied by Wust *et al.* (1982) such as, plasmid analysis and soluble protein polyacrylamide gel electrophoresis (PAGE). Most of the isolates were indistinguishable by these techniques which indicated that cross infection had taken place. Sell *et al.* (1983) used a combination of bacteriocin and bacteriophage typing methods and was also unsuccessful. Nakamura *et al.* (1981) then used serum agglutination as typing method which was later improved as a standard serotyping method to compare with other typing techniques. All these techniques were developed to understand the epidemiology of *C. difficile* infection at the local level. Since these methods indicated that cross-infection had occurred and they were investigated to study the epidemiology in Outbreaks. For this purpose comparison between typing schemes were performed and a good correlation between the types recognised by plasmid profiling, serotyping and PAGE were found by Mulligan *et al.* (1988).

Whole cell fingerprinting by pyrolysis mass spectrometry (PMS) was successfully used as a means of investigating putative *C. difficile* outbreaks (Magee *et al.,* 1993). This method has the advantage that it can cope with a large throughput of isolates and has a high degree of discrimination (Brazier 2001). However apart from its initial cost of the equipment inter laboratory reproducibility was poor.

Molecular typing methods are generally regarded as superior to phenotypic methods in terms of the stability of expression and greater degrees of typeability, and a number of methods have been applied to *C. difficile* (Brazier 2001). Plasmid profiling proved largely unsuccessful due to the sparse distribution of these extra chromosomal genetic elements within the species. Kuijper *et al.* (1987) used whole cell DNA restriction endonuclease analysis (REA) using

HindIII for *C. difficile* chromosomal DNA analysis. REA is highly discriminatory and a reproducible method although technically demanding and a labour intensive method. Restriction fragment length polymorphisms (RFLP) is an alternative genotypic method that involves initial REA digestion followed by gel electrophoresis and southern blotting with selected labelled nucleic acid probes to highlight specific restriction site heterogeneity. This method was first applied to *C. difficile* by Bowman *et al.* (1991). However this method is also labour intensive and these methods were surpassed by techniques based on polymerase chain reaction (PCR) (Brazier and Borriello 2000).

Arbitrarily primed PCR (AP-PCR) is a genotypic method that permits the detection of polymorphism within the target genome without prior knowledge of the target nucleotide sequence. This was viewed as a potentially useful method for *C. difficile* by McMillin and Muldrow (1992) and others. But results showed poor reproducibility. PCR ribotyping uses specific primers complimentary to sites within the RNA operon and was first applied to *C. difficile* by Gurtler (1993) who targeted the amplification process at the spacer region between the 16S and 23S rRNA regions. This part of the genome was shown to be very heterogeneous and *C. difficile* was shown to posses ten copies of the rRNA genes in its genome which varied between strains and as well as different copies of the same genome (Brazier 2001). Because of its discriminatory power, this method has been used routinely by the UK Anaerobe Reference Unit in Cardiff, which has provided a *C. difficile* typing service for the UK since 1995. A library containing 116 distinct ribotypes has been constructed using over 3000 isolates (Stubbs *et al.,* 1999).

Another method, pulsed field gel electrophoresis (PFGE) allows the whole genome to be analysed after digestion with rare cutting restriction endonucleases, such as *SmaI, KspI, SacII or NruI.* Although analysis and comparison between PFGE gels are simple and discriminatory, disadvantages include the cost of the equipment, slowness of the electrophoresis procedure and its complexity. A comparison of the three methods PCR ribotyping, AP-PCR and PFGE by Bidet *et al.* (2000) concluded that PCR ribotyping, although marginally less discriminatory than PFGE offered the best combinations of advantages and is the method most commonly used today.

1.2.6 Variable Number Tandem Repeats (VNTR)

The analysis of DNA has been used in a large number of studies in bacterial taxonomy, bacterial typing and for further understanding the basic mechanisms of evolution (Gurtler and Mayall 2001). A wide range of molecular techniques including PCR ribotyping, Pulsefield Gel Electrophoresis (PFGE), Amplified fragment length polymorphism (AFLP), Multi locus sequence typing (MLST) and Random Amplified Polymorphic DNA (RAPD) are used as typing methods for the characterisation and differentiation of bacterial species down to the strain level. Strain differentiation has become an invaluable tool in epidemiological investigations and the generated data could be used for comparison between laboratories, regions and countries (Kanduma *et al.,* 2003). In addition, strain identification can be useful to determine the outbreak strain and distinguishing it from epidemiologically unrelated isolates (Ramazanzadeh and McNerney 2007).

Tandemly repeated sequences were first described in eukaryotic genomes many years ago (van Belkum *et al.,* 1998) and these were thought to occur in several to thousands of copies, dispersed throughout the genome (Jeffreys *et al.,* 1985). These sequenced elements showed hypervariability among individual persons and were used to prepare DNA fingerprints that are specific to each individual (Jeffreys *et al.,* 1985). Loci with short sequence repeats (SSRs) of one–three bp are generally referred to as microsatellites, whilst loci with 10–100 bp are referred to as minisatellites (Haddad *et al.,* 2004). The repeats tend to vary in size, location and complexity (Lindstedt 2005). With the sequencing of bacterial genomes, it became evident that micro-organisms also contain a high number of these direct repeats. Since many of these loci show hypervariability in their repeat numbers in humans and animals, they are also referred to as Variable Number Tandem Repeat (VNTR) loci (Ramazanzadeh and McNerney 2007). Variability observed in VNTRs is thought to be caused by a mechanism called slipped-strand mispairing (SSM) (Strand *et al.,* 1993), where DNA repeats are inserted or deleted during DNA duplication (Figure 1.3).

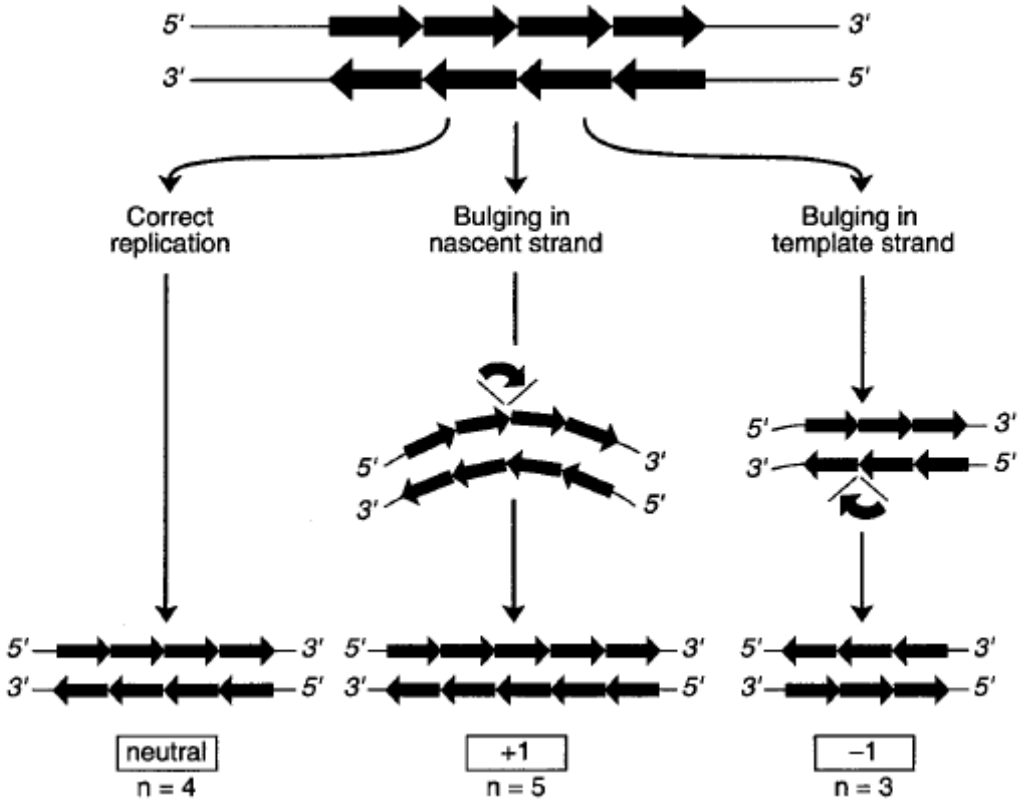


Figure 1.3. Schematic representation of the mechanism of SSM during replication, which results in shortening or lengthening of SSRs. Individual repeat units are identified by arrows; bulging is the presence of non-base-pair base residues interrupting a regular 2-strand DNA helix. Bulging in the nascent strand leads to a larger number of repeat units; bulging in the template strand results in a smaller numbers of units. During replication, bulges can occur in both strands, and the effect of insertion or deletion can be neutralized by occurrence of the adverse event. The number of repeat units can decrease or increase by multiple repeats once multiple bulging in one strand has occurred (Adapted from van Belkum *et al.,* 1998).

VNTRs have been studied extensively in eukaryotic genomes and have been successfully used for identification (van Belkum *et al.,* 1998). The first bacterial species in which they were identified was *Mycobacterium tuberculosis* (Supply *et al.*, 2000) and is described as Mycobacterial Interspersed Repeat Units (MIRUs). Recently, a number of studies have used VNTRs as a means for identification of bacterial isolates i.e. *Bacillus anthracis* (Lista *et al.,* 2006), *Yersinia pestis* (Klevytska *et al.,* 2001), *Haemophilus influenzae* (van Belkum *et al.,* 1997), *Vibrio cholerae* (Ghosh *et al.,* 2008), *S. aureus* (Sabat *et al.,* 2003, Hardy *et al.,* 2004) and many other bacteria. Although most of the early work on tandem repeats was used for

strain identification of bio terror-related micro-organisms, it is now being used for human pathogens such as *Salmonella* serovars (Lindstedt *et al.,* 2003) and *E. coli* (Noller *et al.,* 2006). Many studies have named the repetitive DNA structure differently; eg. MLVA, MIRU, and SIRU, but in this study they will be referred to as VNTRs.

Due to the rapidly increasing knowledge of genome sequences, the search for faster, less labour intensive and low cost methods has lead to DNA-based typing methods. Since VNTR assays are based on PCR amplification of specific locus where targets are known, it has become a promising tool in strain typing (Lindstedt 2005, van Belkum 2007 and Budowle *et al.,* 2005). Also, the ability to use fluorescent dyes allows multiplex assays to be developed, giving amplicons different colours and separating it on capillary sequencers, making it a rapid, reliable, high-throughput strain identification technique. The typing methodology involves PCR amplification of the entire tandem repeat loci using primers positioned in the flanking regions of the locus.

1.2.7 Treatment and infection control

Antibiotics remain the most frequently used treatment for CDAD while a number of nonantibiotic agents are under investigation for treatment. For several years oral vancomycin was thought to be the treatment of choice, but later studies revealed that oral metronidazole was therapeutically equivalent to oral vancomycin for treatment for CDAD (Teasley *et al.*, 1983). The main advantages of using metronidazole were, it was less expensive and was less likely to promote the spread of vancomycin-resistant enterococci (Cloud and Kelly 2007). It was also suggested that the patients with severe symptoms who cannot receive oral treatment, a combination of intravenous vancomycine and intravenous metronidazole should be considered (Stoddart and Wilcox 2002).

Although metronidazole and vancomycin remains the main treatment for CDAD, several new antibiotics such as rifaximin and nitazoxanide are also under investigation and are already approved for other gastrointestinal infections (Blossom and McDonald 2007). Also tolevamer, a *C. difficile* toxin-binding resin, monoclonal antibodies directed at toxin A and B and a *C. difficile* vaccine for immunisation are in early stages of development (Wilcox 2003). Since perturbation of the bowel flora by antibiotic use, it was thought that probiotics maybe effective in prevention of CDAD or in treatment of recurrent CDAD. The most common probiotics include *Lactobacillus rhamnosus* GG, *Saccharomyces boulardii* and other probiotic mixtures including bifidobacteria (Cloud and Kelly 2007). Although a number of studies indicated mixed results for probiotic use, it as been shown recently that in a mouse model, intestinal immunoglobulin A and antitoxin A secretion is stimulated by *S. boulardii* indicating its possible efficacy (McFarland 2006).

Health care facilities remain the main centre for *C. difficile* transmission and measures should be taken to prevent the infection control i.e. patient to patient transmission. When a patient is diagnosed with CDAD, prevention methods such as isolating the patient, patients with CDAD should be allowed to use the same bathroom and once the diarrhoea ceases, transferring the patients to another room should be taken. Also healthcare workers should always take measures such as wearing gloves and gowns for all patient contacts (Blossom and McDonald 2007). Hand hygiene also plays a major role in infection control. Since *C. difficile* spores have the ability to survive under any environmental conditions, measures should be taken to prevent

the spread through spores. It was shown that hand washing by alcohol based sanitizes are unable to prevent the spread. Hence, healthcare workers were adviced to wash their hands with soap and water after removal of gloves. Also a special attention should be given to environmental cleaning of care areas those accommodate patients with CDAD. At present the most commonly used sporicidals i.e. household bleach, contains 5000 ppm of sodium hypochlorite and fresh solutions should be used to clean the surfaces of hospitals in order to prevent the infection spread through spores (Blossom and McDonald 2007).

1.3 Genomics to Proteomics

With the completion of the human genome project and the sequencing of many prokaryote and eukaryote genomes, interest in studying proteins has risen to determine the identity and quantity of expressed proteins (Graves and Haystead 2002). While the genome remains stable to a large extent, the proteome of a cell reflects the immediate environment in which it is studied. In response to internal or external cues, proteins can be modified by post-translational modifications (PTM), undergo translocations within the cell, or be synthesized or degraded (Mirza and Olivier 2008).

Although genomics provide a vast amount of information and also DNA/ RNA is relatively easy to work with, the information on gene regulation and protein identification cannot be obtained from DNA/ RNA analysis. After synthesis on ribosome, simple chemical groups or complex molecules may be attached to the proteins and following translation proteins are chemically change through PTM, mainly through the addition of carbohydrate or phosphate groups. Hence one cell can contain between one and more than 100,000 copies of a single protein (Celis and Gromov 1999). Therefore DNA sequence analysis does not predict the active form of a protein and RNA quantitation does not always reflect the corresponding protein levels. Thus, examination of the proteome of a cell is like taking a "snapshot" of the protein environment at any given time.

Considering all the possibilities, it is likely that any given genome can potentially give rise to an infinite number of proteomes and is far more complex than the genome (Cho 2007). Hence DNA sequence analysis alone cannot provide an accurate profile of protein abundance, structure and activity. Also, proteins are directly involved in both normal and disease associated biochemical processes, a more complete understanding of the diseases maybe gained from looking directly into the proteins. The term "proteomics" was first coined in 1995 and was defined as the large-scale characterisation of the entire protein complement of a cell line, tissue, or organism (Anderson and Anderson 1996; Wilkins *et al.,* 1995 and Tyers and Mann 2003). However the first protein studies that can be referred to as "proteomics" began much earlier in 1975 with the introduction of the two-dimensional gel (2-DE) by O'Farrell (O'Farrell 1975).

Due to the complex nature of the proteome, proteomic analysis has become a challenging field. The proteomics field can be divided into two main areas; expression proteomics and functional proteomics. The former is the study of differentially expressed proteins in various conditions while the latter refers to the study of proteins involved in function such as signalling pathways etc. (Graham *et al.,* 2007). A vital role of proteomics is the discovery of disease biology, mechanisms and new drug targets (biomarkers) which will ultimately help design products to prevent, diagnose and treat diseases (Mirza and Olivier 2008). Biomarkers can be broadly defined as indicators of normal biological or pathological disease states as well as pharmacological responses to therapeutic inventions (Elrik *et al.,* 2006). Hence identification of earlier and specific biomarkers should provide target molecules to diagnose disease and improve in patient treatment.

Earlier proteome analysis mainly relied on gel-based proteomics such as SDS-PAGE and 2-DE that permits the separation of thousands of proteins based on mass and charge. High resolution 2D-PAGE can resolve up to 10,000 protein spots per gel and could be visualised using stains such as Coomassie blue, silver, SYPRO Ruby and Deep Purple (Lauber *et al.,* 2001). The limitations of 2D-PAGE mainly are that it is a time consuming and labour-intensive process has low dynamic range and inherent gel-to-gel variability. However, with the rapid progress in mass spectrometry (MS) in the last decade has made it a key technique for the investigation of the proteome and has increasingly become the method of choice for analysis of complex protein samples (Aebersold and Mann 2003; Tyers and Mann 2003; Kolker *et al.,* 2006 and Domon and Aebersold 2006). Due to the high accuracy of this method, which detects peptides up to femtomole to attomole level, it is now possible to detect the proteins by using public protein databases such as NCBI (Hirsch *et al.,* 2004) which could be easily accessed electronically.

In order to analyse a complex protein sample by MS, involves an initial fragmentation step, a separation step for the fragmented complex protein mixture and mass spectral analysis of the fragmented particles. The fragmentation is normally done by digesting the protein sample using trypsin and separating the fragmented protein mixture using chromatography (LC) and analysed using tandem MS (MS/MS) (Kolker *et al.,* 2006). Subsequently, an alternative to 2-DE, a gel-free approach was developed as multidimensional protein identification technology (MUDPIT) to overcome the limitations of 2D-PAGE (Roe and Griffin 2006). Depending on

the methodology used to identify certain proteins, two strategies have been introduced (Mirza and Olivier 2008 and Graham *et al.,* 2007); top down and bottom up proteomics. The top down methodology identifies the proteins/ peptides without the prior proteolytic digestion of the sample while bottom up methodology analyse the peptides derived from the digested samples (Zhou and Veenstra 2008, Lin *et al.,* 2003).

In addition to protein identification, it is essential to quantify the proteins in order to understand its role in the organism (Bantscheff *et al.,* 2007) and several mass spectral approaches have been established to quantify proteins. Traditionally, gel-based approaches were used in order to quantify the proteins while new methods such as stable isotope labelling and label-free approaches have been introduced (Kito and Ito 2008). Differential gel electrophoresis (DIGE) has been established to overcome the limitations of 2DE as well as to highlight differentially expressed proteins. In DIGE, proteins can be detected with highly sensitive fluorescent dyes known as CyDye fluors (Marouga *et al.,* 2005) and up to three different samples can be identified on a same gel, each labelled with a different fluorescent dye. For the relative quantification of different protein samples, the samples are labelled with stable isotopes and analysed by MS e.g. isotope coded affinity tagging (ICAT) and stable isotope labelling by amino acids (SILAC) (Kito and Ito 2008) while for both relative and absolute quantification a more advanced MS method such as isobaric tagging (iTRAQ) has been used. Recently, label-free methods for protein quantification have been used. But this technique also has its limitations (Kito and Ito 2008).

1.4 Proteomic techniques used in this study

1.4.1 Mass spectrometry

One of the most important developments in protein identification has been the development of mass spectrometry technology. The first major technology to emerge for the identification of proteins was the sequencing of proteins by Edman degradation (Edman 1949). In contrast to Edman sequencing, mass spectrometry is a high sensitivity, high-throughput technique used to acquire both molecular weight and sequence information of both proteins and peptides.

Pyrolysis mass spectrometry (PY-MS) was one of the first MS techniques to be used to fingerprint and discriminate between bacteria at species and even strain level (Goodfellow 1995). But one of the major disadvantages of PY-MS is that little structural information is obtained as the size and range of masses obtained is small. The rapid progress in mass spectrometry in the last decade has made it a key technique for the investigation of the proteome, which was initiated by the development of the two ionisation techniques, Matrix Assisted Laser Desorption/Ionisation (MALDI) (Karas and Hillenkamp 1988) and Electro Spray Ionisation (ESI) (Tanaka *et al.,* 1987) which has led to significant improvements in the central step of protein identification. Due to the high accuracy of this method, detection of peptides in the femtomole to attomole range with an accuracy of <5 parts per million (ppm) (Clauser *et al.,* 1999) is now possible by using "protein databases" such as, National Centre for Biotechnology Information (NCBI).

Thomson developed the first mass spectrometer (then called the parabola spectrograph) during the first decade of the 20th century. There are many different types of mass spectrometers, but the basic principle is universally applicable: the sample is ionised and subsequently the ions are separated according to their mass to charge (m:z) ratio and displayed as a mass spectrum. Hence a mass spectrometer comprises of two major components which are an ionisation source for the production of ions and a mass analyser to detect the masses of produced ions. Throughout the 1980s emerging desorption/ionisation techniques such as Plasma Desorption (PD), Laser Desorption (LD) and Fast Atom Bombardment (FAB) were evaluated for the generation of molecular ions from microorganisms. But there were drawbacks and limitations for these techniques such as ionisation efficiency, mass range of the ionisable compounds and

other practical issues. With the rapid development in MS, during the 1980s-1990s, Matrix Assisted Laser Desorption/Ionisation (MALDI) and Electro Spray Ionisation (ESI) evolved as new methods for the ionisation of biological samples. In both methods, peptides are converted to ions by the addition or loss of one or more protons. ESI and MALDI are “soft” ionization methods that allow the formation of ions without significant loss of sample integrity. This is important because it enables accurate mass information to be obtained for proteins and peptides in their native states (Zhou and Veenstra 2008).

1.4.2 Matrix Assisted Laser Desorption/Ionisation (MALDI)

Matrix Assisted Laser Desorption/Ionisation (MALDI) is a soft ionisation technique used in mass spectrometry, allowing the analysis of biomolecules / biopolymers such as proteins, peptides and sugars and large organic molecules which tend to be fragile and fragment when ionised by more conventional ionisation methods such as electro spray ionisation (ESI). The term Matrix Assisted Laser Desorption/Ionisation was first coined in, 1985 by Franz Hillenkamp, Michael Karas and their colleagues (Karas *et al.,* 1985). They found that the amino acid alanine could be ionised more easily if it was mixed with the amino acid tryptophan and irradiated with a pulsed 266 nm laser. Thus tryptophan regarded as an "absorbing matrix" resulting in molecular ion formation of alanine.

Until recent years, desorption of bioorganic compounds in the mass range above 10,000 Da could be done only by Plasma Desorption Mass Spectrometry (PDMS). In 1988, Tanaka reported the laser desorption of protein molecular ions up to a mass range of 34 000 Da using "ultra fine metal plus liquid matrix method" that combined 30 nm cobalt particles in glycerol with a 337 nm nitrogen laser for ionisation (Tanaka *et al*., 1988). However with the discovery of MALDI, Karas and Hillenkamp were subsequently able to ionise the protein albumin, up to 67,000 Da with a 266 nm laser and a solution of nicotinic acid served as the absorbing matrix (Karas 1988).

While the exact desorption/ionization mechanism for MALDI is not known, it is generally believed that MALDI causes the ionization and transfer of a sample by mixing with solid organic chemical, the matrix. In the MALDI process the peptides or proteins which need to be identified are placed on a target plate, co-crystallised with a matrix solution usually a UV-absorbing weak organic acid and air dried (Karas 1988). Irradiation of this analyte-matrix mixture by a laser results in the vaporization of the matrix, which carries the analyte with it. The matrix plays a key role in this technique. The co-crystallized sample molecules also vaporize, but without having to directly absorb energy from the laser. Ions produced by MALDI are generally singly-charged ions, but doubly charged ions have also been observed *e.g.* in the case of matrix molecules and other stable molecules (Siuzdak 1994).

Once in the gas phase, the desorbed charged molecules are then directed electro statically from the MALDI ionization source into the mass analyzer to perform the mass analysis. Time-of-flight (TOF) mass analyzers are often used to separate the ions according to their mass-to-charge ratio (m/z). It is calculated from the time taken by the ions to travel through the length of the flight tube to the detector (Figure 1.4).

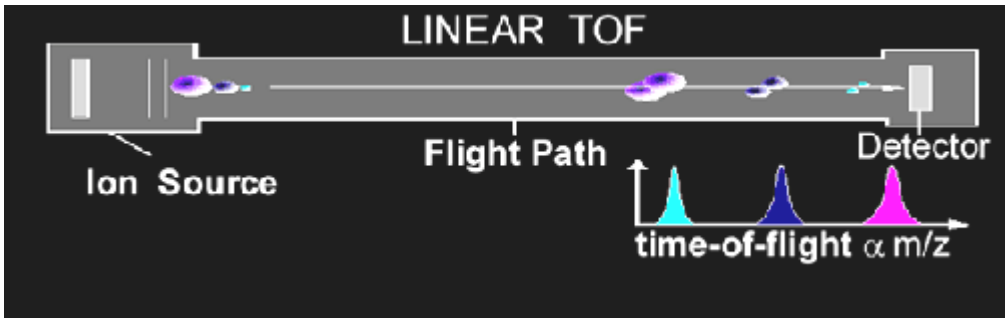


Figure 1.4: Basic components of a Linear-TOF Mass spectrometer. The left side shows the ion source where analytes ionised. The narrow flight tube during which charged ions travel to reach the detection; the smaller masses arriving first (Modified from The Institute of Food Research and John Innes Centre Joint Proteomics Facility).

1.4.3 The Matrix

A non-volatile solid material facilitates the desorption and ionization process by absorbing the laser radiation. As a result, both the matrix and any sample embedded in the matrix are vaporized. The matrix also serves to minimize sample damage from laser radiation by absorbing most of the incident energy.

A variety of matrices have been suggested for the use of MALDI. The selection of a compound for the use as a MALDI matrix must be soluble in a solution with the analyte; if this is not possible the matrix and the analyte must be deposited separately on to the target plate. The matrix consists of crystallized molecules, fairly low molecular weight and acidic, therefore act as a proton source to encourage ionization of the analyte. A further requirement for a matrix is to be chemically inert in terms of reactivity with the analyte (Beavis and Chait 1989).

The three most commonly used matrices are 3, 5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid) (Beavis and Chait 1989), α-cyano-4-hydroxycinnamic acid (alpha-cyano or alpha-matrix) and 2, 5-dihydroxybenzoic acid (DHB). A solution of one of these molecules is made, often in a mixture of highly purified water and an organic solvent normally acetonitrile (ACN) or ethanol. Trifluoroacetic acid (TFA) may also be added.

Matrix Assisted Laser Desorption/Ionisation Time-of-flight mass spectrometry (MALDI-TOF-MS) has a number of advantages over other methods of mass spectrometry when analysing biological material. Crude proteins can be analysed without the need for extensive separation (Hillenkamp *et al.,* 1991) and the insensitivity to impurities can eliminate sample purification steps, reducing the time needed for analysis (Easterling *et al.,* 1998). A variety of studies have shown that MALDI-TOF-MS may be used for the rapid analysis of biological components of bacterial cells (Dai *et al.,* 1999; Chong *et al.,* 1997; Liang *et al.,* 1996). Another great advantage of MALDI-TOF-MS is that intact cells can be taken directly from a colony and analysed within minutes, reducing the sample preparation time (Holland *et al.,* 1996; Claydon *et al.,* 1996; Krishnamurthy and Ross 1996; Welham *et al.,* 1998, Fenselau and Demirev 2001). Intact cell MALDI-TOF-MS, also known as ICM-MS generally detects the surface components which have not been studied systematically up to now. Since many of the interesting properties associated with microbial physiology, virulence and pathogenicity are associated with the surface components of the cell, MALDI-TOF-MS offers the large scale comparative analysis of such molecules and provides the means to study their diversity among the microorganisms.

1.4.4 Waters® (MMU) database

With the aim of establishing a high throughput system for the characterisation of microbes, based on intact cell MALDI-MS (ICM-MS), a database was created by a consortium comprising Waters Inc, Manchester Metropolitan University (MMU) and Health Protection Agency, Centre for Infections (HPA, CFI). It comprised nearly 5000 MALDI spectra of Type and reference stains (Keys *et al.,* 2004) obtained from the National Collection of Type Cultures (NCTC), HPA. The parameters which are likely to affect the reproducibility of the mass spectrum were studied extensively prior to assembling the database. Some of these were, sample preparation, growth phase, culture conditions, sample storage, mass range of ions, reproducibility between instruments and the methodology prior to database entry (Shah *et al.,* 2002). For database entry twelve replicates per sample were analysed on 96-well target plates containing central wells for peptide standards to correct against mass drift during analysis. The quality of the data was assessed statistically prior to database addition using root mean squared values of <3.0 as the criterion for rejection.

To date, MALDI-MS has shown that type and laboratory reference isolates often differ from clinical isolates (Rajakaruna *et al.,* 2009). Therefore it is uncertain how the database will perform when challenged with clinical isolates. A major aim of this study was therefore to collect field isolates from a hospital, give them minimal time to adapt to the laboratory conditions, to ascertain their purity and attempt to identify isolates by comparative analysis of their MALDI profile to an existing profile in the database described above.

1.4.5 Spectral ARchive And Microbial Identifications System (SARAMIS™) database (AnangnosTec, Germany)

During the course of the above study, another database referred to as SARAMIS™ was initiated by AnagnosTec in Germany (Kallow *et al.*, 2006). The SARAMIS™ software and database of human and veterinary pathogens, food pathogens, environmental and natural product producing microorganism began development using MALDI-TOF-MS. The database consisted of, reference spectra and software that allowed the comparison of the mass fingerprint of an unknown sample to the reference spectra (http://www.anagnostec.eu/home.html). Due to the soft ionisation of MALDI, mass signals from 2000- 20,000 Da were collected to assemble the database. It was subsequently found that most of the signals were mainly from stable ribosomal proteins (bacteria) and surface proteins (fungi) (Ryzhov and Fenselau 2001 and Pineda *et al.*, 2003). This was confirmed when the *E. coli* K12 whole cell extracts were analysed and the mass signals obtained superimposed with the *E. coli* K12 ribosomal proteins (Figure 1.5).

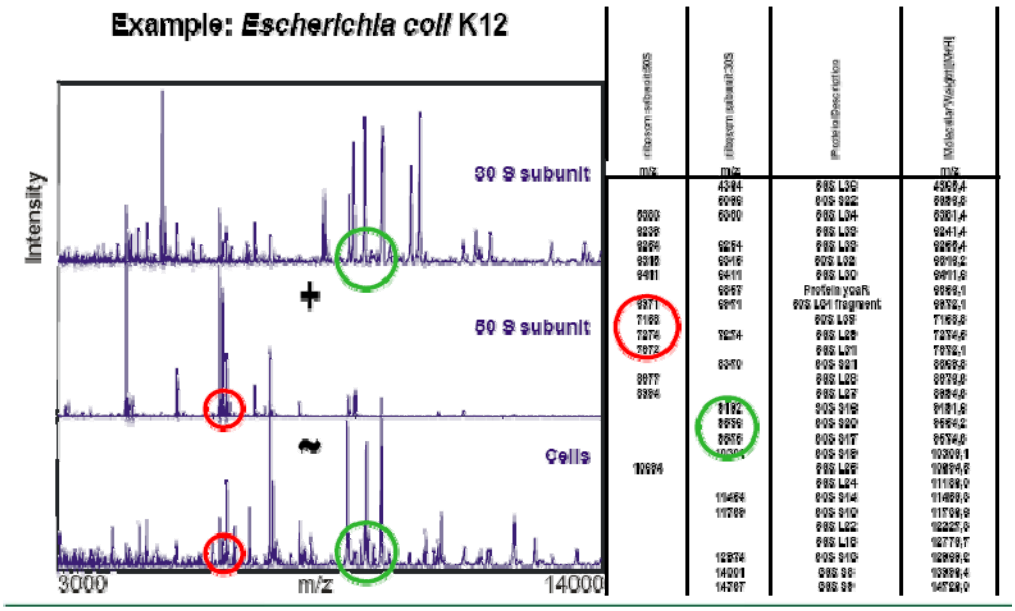


Figure 1.5: Example of *E. coli* K12 when analysed using SARAMIS™ and most of the peaks detected were of stable ribosomal proteins (Taken from Mass Spectrometry for Microbial Proteomics 2010: Wiley).

1.4.6 Quality control, sample preparation and different matrix:

In order to ensure the quality of microorganism identification with SARAMIS™, different procedures and control mechanisms have been used. The SuperSpectra™ have been generated from replicate mass spectra of multiple isolates of a single species or sub-species. To generate the SuperSpectra™, high quality mass spectra of a species of interest were collected from multiple isolates, grown under varying conditions and in different laboratories. Beside the fingerprint spectra, the identity of every isolate was also determined by different methods, e.g. biochemical methods and 16S RNA-sequencing identification procedures. The result of the identification by SARAMIS™ SuperSpectra™ was then compared with the other methods. In addition, the SARAMIS™ microorganism identification database is routinely evaluated by inter laboratory comparisons.

To obtain a successful MALDI-TOF-MS analysis, good sample preparation is a crucial step. AnagnosTec has developed several sample preparation methods using different matrix solutions, which could be used to yield high quality spectra (Table 1.1 and Figure 1.6).

(a). Sample preparation 1 (alpha-cyano-4-hydroxy cinnamic acid)

This sample preparation is recommended when rapid species identification is desired. It requires a lower number of laser shots, thus saving time and maintenance costs. However, in the resulting spectra lower numbers of peaks are recorded compared to the other sample preparation methods (Table 1.1 and Figure 1.6).

(b). Sample preparation 2-(2, 5-Dihydroxybenzoic acid/ DHB)

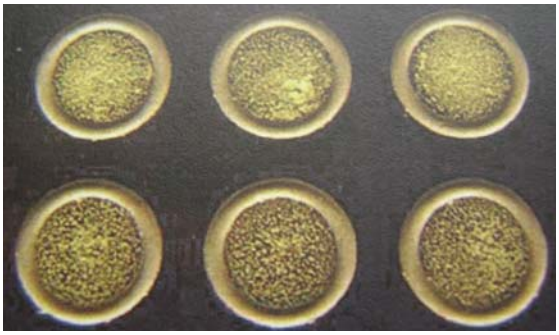
In addition to identification if strain typing is desired, this method is recommended. It requires a higher number of laser shots and as a result, the spectra contain higher number of peaks (Table 1.1 and Figure 1.6).

(c). Sample preparation 3

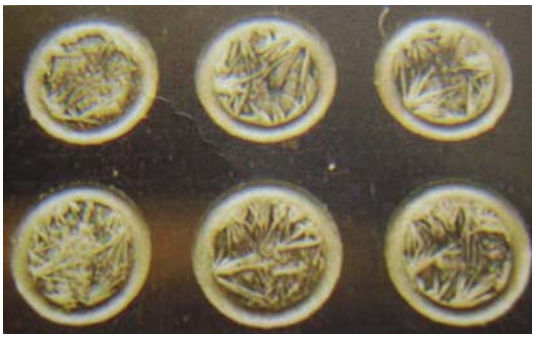
If the spectra obtained from above mentioned procedures are not satisfactory for identification, as an additional step, the sample is pre-extracted using 25 % formic acid. Formic acid increases the permeability of the cell wall and protein extraction efficiency.

Matrix	Species identification	Strain typing
DHB	1 min / 1,500 laser shots	1-2 min / 2,000 laser shots
alpha-cyano-4-hydroxy cinnamic acid	30 s / 400 laser shots	Not suitable

Table 1.1: Overview of the sample preparation procedures using different matrix solutions.



(a)



(b)

Figure 1.6: (a) Samples prepared with Alphacyano matrix. Small spherical crystals are homogeneously distributed on the spots. (b) Samples prepared with DHB matrix. Large, needle shaped crystals are found at the outer border of the spots.

1.4.7 Review of MALDI-TOF-MS used for bacterial identification and characterisation

The first attempts to utilise MALDI-TOF-MS for the characterisation of intact bacterial cells were reported in 1996 by several workers. Holland and co workers (1996) who demonstrated the ability to differentiate isolates representing genus *Pseudomonas* by using characteristic mass ions while Krishnamurthy and Ross (1996) reported species-specific biomarkers for *Bacillus* sp., *Yersinia pestis* and *Francisella tularensis*. Claydon *et al.* (1996) further confirmed the potential of the approach for the rapid identification of gram negative and gram positive bacteria representing different genera and species. In 1998, an improved sample preparation method was introduced by Welham *et al.* (1998), which enabled the identification of species-specific markers in a wider mass range compared to other studies of gram negative bacteria. Anaerobic bacterial species such as *Porphyromonas, Bacteroides* and *Prevotella* were also delineated using MALDI using a simple sample preparation method (Shah *et al.*, 2002).

A study by Shah *et al.* (2000) on gram positives and gram negative bacteria, suggested the potential of a database of archived mass listings of profiles to be utilised in order to compare an unknown to a given spectrum for rapid identification. More recently, Carbonelle *et al.* (2007) reported the separation of coagulase-negative staphylococci from other closely related taxa using an archived spectral database. Rapid discrimination of 24 bacterial species associated with food contaminants were carried out by Mazzeo *et al.* (2006). This work resulted in the development of a database for public access to identify food contaminant microorganisms that cause human diseases. A subsequent study by Barbuddhe *et al.* (2008) developed and expanded MALDI-TOF-MS methods for analysing multiple *Listeria* species.

The above studies were mainly based on identifying biomarkers in lower mass range for differentiation at genus or species level. In 1999, Dai *et al.,* used solvent extracts of *E. coli*, fractionated using HPLC followed by MALDI analysis in order to identify low mass proteins and peptide biomarkers (Dai *et al.,* 1999). Arnold and Reilly (1998) focused on the HPLC separation of ribosomal proteins and identified around 55 distinct components. The ability to differentiate minor differences between phylogenically similar bacteria by MALDI analysis was demonstrated by Lynn *et al.* (1999), who reported family-specific biomarkers within the *Enterobacteriaceae.*

The thrust of MALDI-MS application has been clinical. For example, a method has been described for rapid discrimination of MRSA and MSSA using ICM-MS (Edwards-Jones *et al.,* 2000). The clarity of discrimination between MRSA and MSSA strain spectra and the speed of the method suggests a potential role for ICM-MS in the diagnostic laboratory, which could lead to major improvements in the treatment strategy for infected patients. However, another study by Bernardo *et al.* (2002), who used cell extracts of MRSA and MSSA was unable to discriminate between the two using MALDI but concluded that MALDI could be used for discrimination of clonal isolates of MRSA, which might be useful to track epidemic outbreaks of both community and hospital isolates.

A study by Haag *et al.* (1998) reported the identification and speciation of pathogenic *Haemophilus* isolates. *H. ducreyi* was shown to be distinct from other genera and species and strain differences were also found. Strain specific biomarkers were detected for *H. pylori* (Nilsson 1999) based on MALDI. Reference and environmental isolates of *E. coli, Salmonella* and *Acinetobacter*, isolated from sewage sludge were successfully identified using MALDI by Ruelle *et al.* (2004).

The above examples serve to demonstrate the growing interest of MALDI-TOF-MS in the identification of pathogenic bacteria. This is mainly due to its speed of analysis, minimal sample preparation and biomass required. MALDI provides characteristic mass spectral fingerprints of bacterial species and provides unique insights into bacterial biology and chemistry based on the detection of specific chemicals or chemical changes in response to environmental or other external influences (Lay 2001).

1.5 SDS-PAGE

Taxonomic relationships between organisms can be estimated using variety of characteristics. The majority of bacterial classifications have, in the past, been based on phenotypic or observable characters. But it suffered major drawbacks and limitations such as, reproducibility and was time consuming. In the last two decades, the use of molecular methods for bacterial identification has been employed with considerable success. The composition and order of amino acids which makes up proteins, is determined by the sequence of nucleotides encoded on the DNA of an organism. Thus, the protein pattern reflects the genome of a particular strain. Electrophoresis of a bacterial protein sample under standardised and reproducible conditions therefore produce a protein banding pattern that is characteristic "fingerprint" of a particular strain (Costas 1995).

Polyacrylamide Gel Electrophoresis (PAGE) of proteins has been used increasingly during the past few decades in bacterial classification and identification (Cato *et al.,* 1982). Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE) is a technique which can be used for separation of whole-cell proteins according to their molecular weight and appears to detect broader taxonomic relationships especially at the species and subspecies level. (Vandamme *et al.,* 1998). Sodium dodecyl sulphate (SDS) is an anionic detergent which denatures proteins by "wrapping around" the polypeptide backbone and SDS binds to proteins fairly specifically in a mass ratio of 1.4:1. In so doing, SDS confers a negative charge to the polypeptide chain and disrupts all noncovalent protein bonds causing the macromolecules to unfold (Garfin 1990). By treatment with disulphide-reducing agents such as, 2-mercaptoethanol or dithiothreitol further denatures the proteins. There are two types of buffer systems in electrophoresis, continuous and discontinuous (Laemmli 1970). A continuous system has only a single separating gel and uses the same buffer in the tanks and the gel. In a discontinuous system, a non-restrictive large pore gel, called a stacking gel, is layered on top of a separating gel called a resolving gel. Each gel is made with a different buffer, and the tank buffers are different from the gel buffers. The resolution obtained in a discontinuous system is much greater than that obtained with a continuous system. SDS-PAGE offers the combination of high resolution and good reproducibility and has clear advantages over the earlier used non-denaturing systems (Costas 1995).

SDS-PAGE has been successfully applied for the identification of several bacterial groups. Differentiation and identification of thermophilic campylobacter's (Vandamme *et al.,* 1997), *Campylobacter lari* (Duim *et al.,* 2003) and identification of bacterial isolates in activated sludge were shown by Hantula *et al.* (1991). A recent study by Santos and co workers demonstrated the reliability of the technique by identifying species and subspecies of Staphylococci by SDS-PAGE (Santos *et al.,* 2009). Costas examined the protein patterns of 50 isolates of *Staphylococcus*, 41 of which were MRSA and showed that the technique could be used to study the epidemiology and the evolution MRSA isolates (Costas *et al.,* 1989). A study by Calderon used the technique to asses the similarities between the *Burkholderia cepacia* complex isolates and to identify common protein bands that could be associated with resistance to certain antimicrobial agents (Calderon *et al.,* 2008). Some other studies included the identification of bacterial groups such as *Bacteroides ureolyticus* (Taylor *et al.,* 1987) and *Providencia rettgeri* (Costas *et al.,* 1989).

1.5.1 NuPAGE® electrophoresis system

The NuPAGE® Bis-Tris [Bis (2-hydroxyethyl) imino-tris (hydroxymethyl) methane-HCL] electrophoresis system is a neutral pH and discontinuous SDS-PAGE medium. It uses pre-cast polyacrylamide mini-gels with ready-made buffers, which provide a better band resolution (NuPAGE® Technical guide) and is compatible with subsequent MS analysis. Although the Laemmli system (Laemmli 1970) is the most widely used SDS-PAGE system, the highly alkaline operating pH causes, band distortion, loss of resolution, or artefact bands. The NuPAGE® Gels have overcome this problem and results in sharper band resolution (1-200 kDa) and accurate results (Moos *et al.,* 1998).

In this study a new system, NuPAGE® (Invitrogen™) 1D precast gels were used to compare the whole cell protein profiles of the bacteria and to assess the reproducibility and resolution of the system against MALDI-TOF-MS. The rational for this is that a large number of isolates can be compared relatively easily. NuPAGE® gels are compatible with mass spectrometry analysis, therefore the proteins can be cut out from gels trypsin digested and identified using MS techniques.

1.6 Surface- Enhanced-Laser-Desorption/Ionisation (SELDI)

The advents of ESI and MALDI have extended the application of mass spectrometry to the study of proteins from complex biological systems. However, complex biological material such as blood, sera, plasma, lymph, whole cells, cell lysates, urine and cellular secretion products typically contain thousands of biological molecules as well as organic and inorganic salts which can interfere with direct MS analysis. Thus, significant sample preparation and purification steps are necessary prior to MS investigation. Classical methods of sample purification such as liquid chromatography, electrophoresis, centrifugation and membrane dialysis are often labour intensive and need higher sample volumes. Furthermore, samples may be lost due to precipitation, dilution effects or non-specific binding.

Towards the end of 1993, Hutchens and Yip introduced the concept of Surface Enhanced Laser Desorption/Ionisation (SELDI) which is essentially MALDI-TOF-MS combined with the use of ProteinChip® arrays as a MALDI target plate (Hutchens and Yip 1993). This combined concepts from different technologies: Surface-Enhanced Affinity Capture (SEAC), Surface-Enhanced Neat Desorption (SEND) and Surface-Enhanced Photo Labile Attachment and Release (SEPAR) (Merchant and Weinberger 2000).

SEND is a process by which analytes, even those of large molecular weight, may be desorbed and ionised without the need for addition of the matrix while SEPAR is a hybrid of SEAC and SEND, where the affinity capture device also functions as an energy absorbing molecule promoting analyte desorption and ionisation. To date, SEAC application show a more similar approach to the SELDI technology. In SEAC, the probe surface plays an active role in the extraction, presentation structural modification or amplification of the sample (Merchant and Weinberger 2000).

Surface Enhanced Laser Desorption/Ionisation (SELDI) ProteinChip® technology has the ability to provide a rapid protein expression profile from a variety of biological samples, which will allow for the rapid characterisation of the microbes in order to assess the heterogeneity in their expression patterns (Schmid *et al.,* 2005). However, unlike the MALDI target surfaces, the SELDI protein chip surfaces are uniquely designed to retain proteins of interest from complex mixtures according to their specific properties. A few µl of a sample of

interest is placed on the chromatographic surface, incubated and washed with appropriate buffer. The proteins of interest are captured on the chromatographic surface and co-crystallised with a matrix solution and analysed by a TOF-MS. The result is a mass spectrum which comprises the mass to charge (*m/z*) values and the intensities of bound proteins/peptides (Issaq *et al.,* 2002 and Graham *et al.,* 2007).

The commercial version of SELDI is the, ProteinChip® array system (Ciphergen Bio systems/ Bio-rad, USA). The current platform consists of linear mode Time-of-Flight mass spectrometer and ProteinChip® arrays with a variety of affinity surfaces and analysis software packages. The protein arrays can be categorised by the nature of their affinity surfaces (Figure 1.7). Currently, there are two types of surfaces available which are chromatographic and biological surfaces (Fung *et al.,* 2001). The chromatographic surfaces include hydrophobic, hydrophilic, weak cation exchange, strong anion exchange and an immobilised metal affinity surface. These surfaces were developed to selectively capture proteins via charge, hydrophobic or metal chelate interactions. The biological surfaces are designed to covalently bind to biomolecules and enable specific protein-protein or DNA-protein interactions such as antibody-antigen, receptor-ligand or nucleic acid-binding protein interactions. The system is most effective at profiling low molecular weight proteins that are < 20 kDa. However, SELDI-TOF-MS is more sensitive and requires smaller amounts of sample compared to 2-D-PAGE (Issaq *et al.,* 2002).

The ProteinChip® platform has been shown as potential screening tool for lung cancer (Yang *et al.,* 2005). It has also proven to be useful in the discovery of potential diagnostic markers for prostate, bladder (Vlahou *et al.,* 2001), breast (Wulfkuhle *et al.,* 2001), ovarian cancers, renal carcinoma (Xu *et al.,* 2009) Alzheimer's disease (Austen *et al.,* 2000), Human tumors (Ball *et al.,* 2002) and profiling of urinary proteins to assess renal function (Hampel *et al.,* 2001). ProteinChip® Arrays have also been used to characterise protein-protein interactions (Stoica *et al.,* 2001) and to detect virulance factors of *Yersinia pestis* (Thulasiraman *et al.,* 2001).

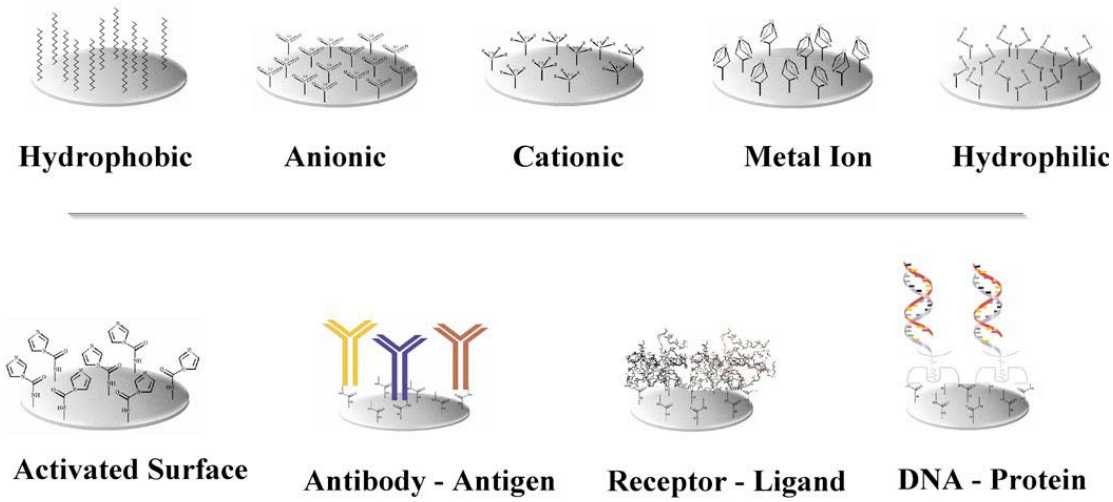


Figure 1.7: Diagrammatic representation of different types of ProteinChips®. (Modified from Ciphergen Bio systems).

1.6.1 Different ProteinChip® arrays generally used to study microorganisms

a) Weak Cation Exchange ProteinChip® array (WCX/CM10)

This surface contains carboxylate groups that interact with positively charged amino acids such as lysine, arginine or histidine at a given pH, on the surface of the analyte.

b) Strong Anionic Exchange ProteinChip® array (SAX2/ Q10)

The active surface of the array contains positively charged, quaternary ammonium groups that interact with negative charges on the surface of target proteins *e.g.* aspartic acid or glutamic acid.

c) Hydrophobic ProteinChip® array (H50)

The active surface of this array contains 16 methylene groups that bind proteins through reverse phase chemistry. The binding is performed to proteins rich in alanine, valine, leaking, isoleucine, phenylalanine, tryptophan and tyrosine.

1.7 Analysis of MS data

Due to the high sensitivity, accuracy, detection limits, speed and diversity of applications, MS has become an important tool in proteomics (Aebersold and Mann 2003). To date, one of the main applications of MS is for biomarker discovery; hence the quality of mass spectra is a key factor before the detection of biomarkers. One of the major problems involved in MS data mining is the complex nature of the data produced i.e. noisy and fuzzy data (Lancashire *et al.,* 2009). In order to overcome these problems, robust computational algorithms have been developed to help the data analysis i.e. cluster analysis, decision trees, principle component analysis (PCA) and machine learning algorithms such as support vector machines (SVM), Artificial neural networks (ANN) (Ball *et al.,* 2002; MS Book chapter 17; 2010; Wiley in press). These algorithms may be divided into two types i.e. supervised and unsupervised learning. Supervised learning is more powerful and based on prior knowledge whereas unsupervised learning needs no extra knowledge about the samples (Mendes 2002).

The fist step in MS data analysis is to minimise the complexity of the datasets produced to improve the quality of the output. This could be obtained at the beginning of sample processing by reducing the contaminants. The next step involves the pre-processing of the samples by the removal of noisy, fuzzy data from the MS which could appear due chemical contaminants present in the samples and the electrical noise which could occur during the processing of the samples in the mass spectrometer. Once the MS are obtained, baseline correction and normalisation of the peaks is carried out in order to subtract the low frequencies from the spectra and to correct experimental variations from the spectra respectively. These steps involve the use of various statistical analytical methods (MS Book 2010: chapter 17: Wiley and Hilario *et al.,* 2004).

Once the peaks have been detected in the mass spectrometer and the mass spectra produced, the next step is to identify the peptides by matching the experimental spectra with the theoretical spectra using database search algorithms. To generate highly sensitive and specific classification models, selection of the features that are able to discriminate between correct and incorrect peptides with high accuracy are necessary. If this is not considered appropriately, the data can lead to false representation of samples, false detection of biomarkers or determination of biomarkers that actually represent noise in the sample.

- **Clustering**

Clustering techniques are an example of an unsupervised learning. One very common method is known as Hierarchical clustering. It functions by arranging the profiles of samples into a tree-like structure so that the most similar profiles lie close together and profiles very different to one-another lie farther apart, allowing for the rapid visual assessment of patterns within the data. The methodology is based on the construction of a distance matrix which enables the two samples with the most similar profiles to be determined. These are then placed together in the tree to form a cluster, and the distance between this newly defined cluster and the remaining sample is calculated. A new cluster is then determined and this process is repeated until all of the samples have been placed into a dendogram. One common use of hierarchical clustering is to create phylogenetic trees or identify structures in populations i.e. Lancashire *et al.* (2005) and Seibold *et al.* (2007) used both hierarchical clustering and PCA to differentiate *Neisseria meningitides* and *Francisella tularensis*.

- **Decision trees**

Decision trees are essentially an extension of clustering approaches where rules are applied to the clustering dendrogram to try and separate individuals in the population into meaningful classes. Adam *et al.* (2001) attempted to identify biomarkers for prostate cancer discovery. They used proteins derived from ProteinChip® surfaces using laser desorption/ionization (SELDI) mass spectrometry and decision trees to discriminate prostate cancer proteins and benign prostate hyperplasia/healthy men.

- **PCA**

One of the common methods used for visualisation and dimensionality reduction in mass spectrometry datasets is Principal components analysis (PCA) (Harrington *et al.,* 2005)**.** This technique appears widely as a software tool for mass spectrometry and provides a useful approach for the visualisation of structures within a population of mass spectrometry runs and can be used to identify the main component ions associated with that structure. PCA transforms the input space into a new space described by what are known as principal

components, which are expressed as linear combinations of the original variables, essentially drawing a line through the data that explains most variation.

- **Support Vector Machines**

Support Vector Machines (SVMs) are a relatively powerful supervised learning algorithm that has been widely implemented in mass spectrometry. It's a binary classification algorithm that separates the positive and negative examples by constructing a straight line or hyper plane that is rotated in multi-dimensional space.

- **Artificial Neural Networks (ANNs)**

Artificial Neural Networks (ANNs) are a form of Artificial Intelligence (AI) capable of modelling for complex systems and fast emerging as one of the most popular tools for complex data analysis. It is a mathematical/computational model based on biological neural networks that organises and process information (Lancashire *et al.,* 2005). ANNs gathers knowledge by detecting patterns and relationships in data and learn through experience, not from programming. The most important aspect of the ANN function is that it is built from a potentially large number of interconnected processing elements, also known as nodes, which work together in a network. Once trained, ANNs can be used to predict the class of an unknown sample of interest. ANNs were first proposed for use in the identification of biomarkers from SELDI- MS data by Ball *et al.* (2002).

1.8 Aims and objectives of this study

The emergence of microbes with low genomic diversity but increasing pathogencity or antimicrobial resistance necessitates the development of high-resolution protein- based techniques capable of exploring minute variation between related isolates. Today, as a result of the rapid and relentless improvements in the analytical capabilities of the technologies, proteomics has changed the way in which many research investigations including physiological function, pathogenesis and diagnostics are approached. In general most proteomic studies commence with gel-based approaches and use MS-based techniques to identify unique biomarker molecules. This study focused on the two major nosocomial pathogens that have been reported to be very diverse species. Here the aim was to develop and assess the potential of various proteomic approaches for the characterisation of these bacteria at the species level and further investigate the intraspecies diversity of one pathogen.

In summary these are:

1). To develop novel protein expression-based methods for the characterisation of various isolates, including both type and clinical isolates of *Staphylococcus aureus* and *Clostridium difficile* using Matrix Assisted Laser Desorption/Ionisation Time of Flight Mass Spectrometry (MALDI-TOF-MS).

2). To characterise MRSA, MSSA and *C. difficile* by their unique intracellular protein biomarkers using Surface Enhanced Laser Desorption/Ionisation (SELDI) ProteinChip® technology and SDS-PAGE.

3). To detect novel loci to differentiate different ribotypes i.e. 027, 001 and other ribotypes of *C. difficile* by using Variable Number Tandem Repeats (VNTR).

Chapter 2

Materials and Methods

2. Materials and Methods

2.1 *S. aureus* Sample Collection - London Hospital isolates for MALDI-TOF-MS study

Samples were collected from the Microbiology Department of the London Hospital following 5 visits made during September and October 2005. Isolates were provided by the Laboratory Manager at each visit, that were representative of the work going through the department at the time.

Each isolate was identified by standard methods used at the London Hospital. The London Hospital reference, culture conditions, Gram stain, and the morphology were recorded for each isolate together with the identification. The cultures were assigned an HPA number and sub cultured to Blood agar slopes for transport to HPA where they were incubated for 24 h at 37 °C. Isolates were plated for purity onto Columbia Blood agar, incubated for 24 h at 37 °C and stored by the Micro bank preservation system (Pro-Lab Diagnostics, UK) at -80 °C. A total of 107 (n = 107) *S. aureus* isolates were collected. Reference isolates/ Type isolates were obtained from the department of National Collection of Type Cultures (NCTC), HPA, London (Table 2.1).

2.1.1 Staphylococcal Reference Laboratory isolates for MALDI-TOF-MS study

Thirty nine isolates were provided by the HPA's Staphylococcal Reference Unit and included cultures that were sent to the laboratory from geographically diverse centres throughout England and Wales for identification and characterisation. All except two were MRSA isolates, isolated during 2007 and included sub-types of current epidemic MRSA isolates (EMRSA-15, -16 and -17) in addition to representatives of three pandemic Community-Associated MRSA lineages (so-called USA300, South West pacific and European clones). The isolates were subcultured on to Columbia blood agar plates for analysis on MALDI-TOF-MS (Table 2.1).

HPA No.	Hospital ID.	HPA No.	Hospital ID.	HPA Sample ID	Source
HPA 26	MRSA	HPA 294	*S. aureus*	H072660527	Clinical isolate
HPA 30	MRSA	HPA 299	*S. aureus*	H072580475	Clinical isolate
HPA 39	MRSA	HPA 318	*S. aureus*	H071920422	Clinical isolate
HPA 40	MRSA	HPA 319	*S. aureus*	H071200363	Clinical isolate
HPA 41	MRSA	HPA 322	*S. aureus*	H072940574	Clinical isolate
HPA 75	MRSA	HPA 323	*S. aureus*	H073020460	Clinical isolate
HPA 76	MRSA	HPA 324	*S. aureus*	H072660333	Clinical isolate
HPA 77	MRSA	HPA 334	*S. aureus*	H072800374	Clinical isolate
HPA 78	MRSA	HPA 339	*S. aureus*	H072340414	Clinical isolate
HPA 79	MRSA	HPA 344	*S. aureus*	H072640555	Clinical isolate
HPA 80	MRSA	HPA 345	*S. aureus*	H072500404	Clinical isolate
HPA 81	MRSA	HPA 346	MRSA	H072240800	Clinical isolate
HPA 82	MRSA	HPA 347	MRSA	H072900348	Clinical isolate
HPA 83	MRSA	HPA 348	*S. aureus*	H072580470	Clinical isolate
HPA 84	MRSA	HPA 353	MRSA	H064560442	Clinical isolate
HPA 85	MRSA	HPA 356	*S. aureus*	H072860414	Clinical isolate
HPA 86	MRSA	HPA 358	*S. aureus*	H041940150	Clinical isolate
HPA 133	MRSA	HPA 389	*S. aureus*	H072680434	Clinical isolate
HPA 134	MRSA	HPA 401	MRSA	H072640554	Clinical isolate
HPA 139	MRSA	HPA 404	*S. aureus*	H072280533	Clinical isolate
HPA 140	MRSA	HPA 405	*S. aureus*	H072820462	Clinical isolate
HPA 144	MRSA	HPA 408	MRSA	H072860467	Clinical isolate
HPA 160	MRSA	HPA 409	*S. aureus*	H072580469	Clinical isolate
HPA 222	MRSA	HPA 410	*S. aureus*	H072820481	Clinical isolate
HPA 229	MRSA	HPA 412	MRSA	H072740472	Clinical isolate
HPA 230	MRSA	HPA 441	*S. aureus*	H072240582	Clinical isolate
HPA 233	*S. aureus*	HPA 442	*S. aureus*	H072820446	Clinical isolate
HPA 239	*S. aureus*	HPA 443	*S. aureus*	H072720523	Clinical isolate
HPA 242	MRSA	HPA 444	MRSA	H072680446	Clinical isolate
HPA 247	MRSA	HPA 489	*S. aureus*	H072660541	Clinical isolate
HPA 248	*S. aureus*	HPA 496	*S. aureus*	H072860465	Clinical isolate
HPA 249	*S. aureus*	HPA 497	*S. aureus*	H072720526	Clinical isolate
HPA 250	*S. aureus*	HPA 499	*S. aureus*	H072820480	Clinical isolate
HPA 256	*S. aureus*	HPA 500	*S. aureus*	RH070000211	Clinical isolate
HPA 257	*S. aureus*	HPA 501	*S. aureus*	RH070000253	Clinical isolate
HPA 258	MRSA	HPA 523	MRSA	H072920482	Clinical isolate
HPA 259	*S. aureus*	HPA 524	MRSA	H072660333	Clinical isolate
HPA 260	*S. aureus*	HPA 545	MRSA	H072820464	Clinical isolate
HPA 261	MRSA	HPA 546	*S. aureus*	H071380629	Clinical isolate
HPA 262	MRSA	HPA 547	*S. aureus*		
HPA 273	*S. aureus*	HPA 549	*S. aureus*		
HPA 279	*S. aureus*	HPA 550	*S. aureus*		
HPA 280	*S. aureus*	HPA 556	*S. aureus*		
HPA 281	*S. aureus*	HPA 563	*S. aureus*		
HPA 284	*S. aureus*	HPA 569	*S. aureus*		
HPA 285	*S. aureus*	HPA 571	*S. aureus*		
HPA 287	*S. aureus*	HPA 573	MRSA		
HPA 293	*S. aureus*				

Table 2.1: *S. aureus* isolates used for MALDI-TOF-MS study. All 'HPA' isolates (95) were collected from Royal London Hospital and the remaining 39 isolates were provided by the Staphylococcal Reference Laboratory at HPA, London.

2.1.2 MRSA and MSSA sample collection for ANN analysis

Most of the isolates used in this part of the study were from a past PhD student (Ines 2008). Out of the 99 isolates, six isolates were from Saunders *et al.* (2004) study and 45 isolates were from the Glasgow collection at HPA, London. Another 48 isolates were obtained from the Staphylococcal Reference Laboratory HPA, London (Table 2.2a and b).

Strain	Date of Isolation	Methicillin susceptibility	Epidemiology
Sanger 476		MSSA	High virulance community acquired
NCTC 8325	1960	MSSA	Laboratory strain
01.1974 W		MSSA	Glasgow collection
00.9302 B		MSSA	Glasgow collection
00.9839 V		MSSA	Glasgow collection
00.11053 Y		MSSA	Glasgow collection
01.1652 H		MSSA	Glasgow collection
01.2463 W		MSSA	Glasgow collection
H355		MSSA	Clinical Isolate
H493		MSSA	Clinical Isolate
H494		MSSA	Clinical Isolate
H495		MSSA	Clinical Isolate
H454		MSSA	Clinical Isolate
H499		MSSA	Clinical Isolate
H512		MSSA	Clinical Isolate
H514		MSSA	Clinical Isolate
H456		MSSA	Clinical Isolate
H440		MSSA	Clinical Isolate
H187		MSSA	Clinical Isolate
H479		MSSA	Clinical Isolate
H108		MSSA	Clinical Isolate
H178		MSSA	Clinical Isolate
H419		MSSA	Clinical Isolate
H450		MSSA	Clinical Isolate
H354		MSSA	Clinical Isolate
H349		MSSA	Clinical Isolate
H420		MSSA	Clinical Isolate
H057		MSSA	Clinical Isolate
H175		MSSA	Clinical Isolate
H278		MSSA	Clinical Isolate
H279		MSSA	Clinical Isolate
H280		MSSA	Clinical Isolate
H281		MSSA	Clinical Isolate
H282		MSSA	Clinical Isolate
H472		MSSA	Clinical Isolate
H248		MSSA	Clinical Isolate
H249		MSSA	Clinical Isolate
H461		MSSA	Clinical Isolate
H096		MSSA	Clinical Isolate
H344		MSSA	Clinical Isolate
H339		MSSA	Clinical Isolate
H328		MSSA	Clinical Isolate
H318		MSSA	Clinical Isolate
H409		MSSA	Clinical Isolate
H595		MSSA	Clinical Isolate
H460		MSSA	Clinical Isolate
H423		MSSA	Clinical Isolate
H401		MSSA	Clinical Isolate
H305		MSSA	Clinical Isolate
H571		MSSA	Clinical Isolate

Table 2.2a: MSSA (50) isolates used for the SELDI-TOF-MS and ANNs analysis. All the isolates were from a previous study (Ines 2008), which was from Saunders *et al.,* 2004 and Glasgow collection.

Strain	Date of Isolation	Methicillin susceptibility	Epidemiology
Sanger 252		MRSA	Hospital acquired
MU 50	1996	VISA*	Hospital acquired
MW 2	1998	MRSA	High virulance community acquired
COL		MRSA	
97.1636 A		MRSA-E15	Glasgow collection
97.1866 D		MRSA-E15	Glasgow collection
98.1859 B		MRSA-E15	Glasgow collection
01.5323 F		MRSA-E15	Glasgow collection
00.9521 M		MRSA-E15	Glasgow collection
98.2028 X		MRSA-E16	Glasgow collection
97.1128 Y		MRSA-E16	Glasgow collection
97.1396 J		MRSA-E16	Glasgow collection
97.2637 D		MRSA-E16	Glasgow collection
00.9523 R		MRSA-E16	Glasgow collection
01.5366 R		MRSA-E16	Glasgow collection
97.2935 K		MRSA	Glasgow collection
00.7895 V		MRSA	Glasgow collection
00.5472 R		MRSA	Glasgow collection
98.1695 K		MRSA	Glasgow collection
01.4964 S		MRSA	Glasgow collection
00.6545 X		MRSA	Glasgow collection
99.3249 A		MRSA	Glasgow collection
03.3200 J		MRSA	Glasgow collection
99.3700 W		MRSA	Glasgow collection
01.6742 M		MRSA	Glasgow collection
01.7633 M-1		MRSA	Glasgow collection
01.1119 S		MRSA	Glasgow collection
01.2426 C		MRSA	Glasgow collection
02.6225 E		MRSA	Glasgow collection
00.10399 P		MRSA	Glasgow collection
00.1924 K		MRSA	Glasgow collection
99.3248 W		MRSA	Glasgow collection
97.1948 S		MRSA	Glasgow collection
99.1133 M		MRSA	Glasgow collection
97.1000 K		MRSA	Glasgow collection
02.3022 X		MRSA	Glasgow collection
01.9694 M		MRSA	Glasgow collection
02.5856 E		MRSA	Glasgow collection
97.3130 D		MRSA	Glasgow collection
02.5099 D		MRSA	Glasgow collection
03.1475 D		MRSA	Glasgow collection
03.1036 Z		MRSA	Glasgow collection
03.7230 R		MRSA	Glasgow collection
H516		MRSA	Clinical Isolate
H527		MRSA	Clinical Isolate
H528		MRSA	Clinical Isolate
H532		MRSA	Clinical Isolate
H610		MRSA	Clinical Isolate
H497		MRSA	Clinical Isolate
***** MRSA displaying intermediate resistance to vancomycin			

Table 2.2b: MRSA (49) isolates used for the SELDI-TOF-MS and ANNs analysis. All the isolates were from a previous study (Ines 2008), which was from Saunders *et al.*, 2004 and Glasgow collection.

2.1.3 *C. difficile* sample collection for MALDI-TOF-MS, SELDI-TOF-MS and SDS-PAGE analysis

One hundred and fifty isolates of *C. difficile*, which span over a 30 year period (1970s- 2007) collected from infected patients, were provided by Dr. I. Poxton at the Microbial Pathogenicity Research Laboratory, MPRL, University of Edinburgh, on ampoules. Isolates were plated for purity on to Columbia Blood agar (CBA) or Fastidious Anaerobic agar (FAA) plates, incubated for 48 h at 37 °C in anaerobic conditions and stored by the Micro bank preservation system at -80 °C. Reference isolates and Type isolates, *Clostridium beijerinckii* (NCTC 13035), *Clostridium chauvoei* (NCTC 8361), *Clostridium bifermentans* (NCTC 6800), *Clostridium difficile* (NCTC 11209), *Clostridium difficile* (NCTC 11207), *Clostridium tertium* (NCTC 2917), *Clostridium septicum* (NCTC 549), *Clostridium butyricum* (NCTC 6084), *Clostridium paraperfringens* (NCTC 10986), *Clostridium perfringes* (NCTC 13112), *Clostridium histolyticum* (NCTC 7124), *Clostridium tetanomorphum* (NCTC 543), *Clostridium sporogenes* (NCTC 275) and *Clostridium putrificum* (NCTC 4718) were obtained from the NCTC, HPA, London and stored by the Micro bank preservation system at -80 °C. The outbreak isolates, B1 (1970s) and T (1980s) were kindly supplied by Professor S. P. Borriello's personal collection. Also the highly virulent strain isolated from the Stoke-Mandeville outbreak in 2006, ribotype 027 (L) were used in the study as well. Another set of isolates were obtained from a hospital outbreak strain collection (Table 2.3).

ID	Original No.	Ribotype	Isolated year
NCTC 13287=R7404		17	-
NCTC 13307=630		12	-
NCTC 13366=R20291		27	-
NCTC 13404		106	-
MPRL 4847	XE123850H	106	-
MPRL 4848	XE123008E	23	-
MPRL 4849	XE118229R	14	-
MPRL 4850	XE117617C	42	-
MPRL 4851	XE118399	1	-
MPRL 4852	XE117124E	1	-
MPRL 4853	121311	2	-
MPRL 4854	120953J	70	-
MPRL 279		Not known	1980
MPRL 296		70	1980
MPRL 307		103	1981
MPRL 369		12	1982
MPRL 402		Not known	1982
MPRL 560		2	1983
MPRL 588		32	1983
MPRL 591		171	1983
MPRL 613		120	1983
MPRL 616		106	1983
MPRL 1037		Not known	1972
MPRL 2783		2	1991
MPRL 808		2	-
B1		5	1970s
L		27	2006
T		1	1980s
MPRL 050		179	1980
MPRL 219		Not known	1980
MPRL 309		12	1981
MPRL 347		Not known	1981
MPRL 368		Not known	1982
MPRL 371		Not known	1982
MPRL 380		5	1982
MPRL 408		Not known	1982
MPRL 558		171	1983
MPRL 589		104	1983
MPRL 604		Not known	1983
MPRL 678		Not known	1984
MPRL 4856		Not known	-
MPRL 4857		Not known	-
MPRL 4858		106	-
MPRL 4859		42	-
N 4(2)		Not known	-
MPRL 407		Not known	-
MPRL 4846	XE123899W	Not known	-
MPRL 002		Not known	1979
MPRL 223		104	1980
MPRL 282		69	1980
MPRL 370		Not known	1982
MPRL 372		12	1982
MPRL 379		12	1982
MPRL 406		Not known	1982
MPRL 665		Not known	1984
MPRL 687		12	1984
MPRL 712		Not known	1984
MPRL 2282		171	1989

Table 2.3: *C. difficile* isolates used for MALDI, SELDI and SDS-PAGE.

ID	Original No.	Ribotype	Isolated year
MPRL 13366		27	-
MPRL 199		Not known	1980
MPRL 202		Not known	1980
MPRL 381		Not known	1982
MPRL 215		Not known	1980
MPRL 248		153	1980
MPRL 250		2	1980
MPRL 255		173	1980
MPRL 291		173	1980
MPRL 844		2	-
MPRL 326		Not known	-
MPRL 059		Not known	-
MPRL 339		23	-
MPRL 349		125	-
MPRL 554		1	-
MPRL 382		33	-
MPRL 556		2	-
MPRL 396		23	-
MPRL 397		Not known	-
MPRL 398		Not known	-
MPRL 405		Not known	-
MPRL 418		Not known	-
MPRL 421		12	-
MPRL 845		140	-
MPRL 422		12	-
MPRL 548		189	-
MPRL 549		1	-
MPRL 602		Not known	-
MPRL 555		Not known	-
MPRL 615		13	-
MPRL 557		2	-
MPRL 559		12	-
MPRL 585		23	-
MPRL 586		1	-
MPRL 587		Not known	-
MPRL 590		165	-
MPRL 592		Not known	-
MPRL 595		Not known	-
MPRL 597		176	-
MPRL 1039		Not known	-
MPRL 842		Not known	-
MPRL 617		Not known	-
MPRL 629		Not known	-
MPRL 688		Not known	-
MPRL 689		Not known	-
MPRL 841		Not known	-
MPRL 308		23	1981
MPRL 319		2	1981
MPRL Outbreak strain		1	-
MPRL Outbreak strain		1	-
MPRL Outbreak strain		1	-
MPRL Outbreak strain		1	-
MPRL Outbreak strain		1	-
MPRL Outbreak strain		1	-
MPRL Outbreak strain		1	-
MPRL Outbreak strain		1	-
MPRL Outbreak strain		1	-
MPRL Outbreak strain		1	-

2.2 MALDI-TOF-MS analysis of *S. aureus* using Waters® (MMU) database

2.2.1 Selection of a suitable growth medium and incubation time for *S. aureus*

In order to find a suitable growth medium and incubation time required for yielding the maximum high density mass spectrum of *S. aureus*, a randomly selected isolate (HPA 30), was maintained on Columbia Blood agar (CBA), Nutrient agar (NA), Chocolate agar (CHOC) and Manitol Salt agar (MSA). The isolate was incubated for different time periods: 24 h, 48 h and 72 h in aerobic conditions at 37 °C. All isolates were subcultured on three successive culture plates prior to MALDI-TOF-MS analysis.

2.2.2 *S. aureus* analysis; Target plate preparation

All reagents were from Sigma UK, unless otherwise stated.

Bacterial cultures were maintained on CBA (Media Department, HPA, London) for 24 h at 37 °C. Chemically cleaned target plates (Waters Corporation, UK) were wiped with methanol (BDH, Poole, UK) and allowed to air-dry prior to use. A small amount of growth was removed from the culture plate using a 1 µl loop and transferred to the target plate. One µl matrix solution containing acetonitrile, water and methanol (BDH, Poole, UK) in a ratio of 1:1:1 (v/v), 0.01M 18-crown-6 ether, 0.1 % formic acid (v/v), saturated with 5-chloro-2-mercaptobenzothiazole (CMBT) at a concentration of 3.0 mg/ml was added to the plate for gram positive organisms. Each plate was calibrated using a peptide mix (Table 2.4), which was prepared by mixing it with alpha-cyano-4-hydroxy-cinnamic acid (α- cyano) at a concentration of 14.0 mg/ml. Matrix solution was sonicated in a sonic bath (Ultrawave, UK) for 10 min prior to use. *Micrococcus lylae* (NCTC 13377), EMRSA (NCTC 13134), MRSA (NCTC 11940), *S. aureus* (NCTC 7727), *S. epidermidis* (NCTC 11047) and *S. epidermidis* (NCTC 11407) were included as 'blind' controls. Twelve replicates were analysed for each sample.

Peptides	*Volume (µl)*	*Final Con.*	*MW of peptides*
Bradykinin	2	1 pmol/µl	1060.2
Angiotensin I	2.6	1 pmol/µl	1296.5
Glu-fibrinogen	3.14	1 pmol/µl	1570.6
Renin	3.52	1 pmol/µl	1759.0
ACTH	4.94	1 pmol/µl	2465.7
Insulin (bovine)	22.93	2 pmol/µl	5733.5
Ubiquitin (bovine)	171.2	10 pmol/µl	8564.9

Table 2.4: List of peptides used to prepare the peptide mix used for calibration with final concentrations and molecular weights.

2.2.3 Data acquisition and processing

Data acquisition was done using a MALDI-TOF Mass Spectrometer (Waters®/ Micromass Ltd. UK) and data processing was performed using MicrobeLynx™ software as described previously (Keys *et al.,* 2004). The instrument was fitted with a 337 nm nitrogen laser. Fifteen spectra per sample well and 10 spectra per lock mass well were collected for each strain in the mass range of 500-10,000 Da. Individual spectral profiles were lock mass corrected with the exact mass of the rennin peak, which is 1759 Da and then the 15 spectral profiles were combined in order to improve the mass accuracy and to produce a reproducible bacterial spectrum for each replicate.

The spectra obtained for the clinical isolates of *S. aureus* were searched against the existing database (MicrobeLynx™ 2005) and the search was based upon an estimation of the probability of the mass spectral peaks in the test spectrum to be comparable with the database spectrum. A list of top eight matches was provided together with Root Mean Square (RMS) value. A high relative and absolute probability and low RMS value indicates a good match (Figure 2.1).

2.2.4 Identification of *S. aureus* by comparative 16S ribosomal RNA sequence analysis

Isolates were grown on CBA (Media Department, HPA) for 24 h at 37 °C. DNA was extracted by resuspending the cells in 60 ml of Prepman Ultra (Applied Biosystems, UK) and heating the suspension for 10 min at 99 °C followed by incubation at 4 °C. The sample was centrifuged at 2500 × g for 10 min and 1 ml was used for PCR. The primers (MWG) used were ANT1F, 5'-AGA GTT TGA TCC TGG CTC AG-3', and 1392R, 5'-ACG GGC GGT GTG TAC AAG-3'giving a product size of 1300 bp. PCR cycle conditions using PCR Ready Mix (Sigma, UK) were as follows: initial denaturation 95 °C for 2 min, followed by 35 cycles of 95 °C for 45 s, 56 °C for 45 s and 72 °C for 60 s. Final extension was carried out at 75 °C for 5 min. Products were cleaned using AMPure PCR Purification Magnetic Beads Kit (Beckman Coulter, UK), following the manufacturer's instructions. The sequencing primers used was 357F, 5'-CTC CTA CGG GAG GCA GCA G-3', and 3R, 5'-GTT GCG CTC GTT GCG GGA CT-3'. DNA sequencing was carried out using Beckman Capillary sequencer (CEQ 8000). The bacteria detected by 16S rDNA PCR were identified by sequence comparison to the GenBank database using BLAST (http://www.ncbi.nlm.nih.gov).

2.2.5 *C. difficile* analysis using the Waters® (MMU) database

Bacterial cultures were maintained on both CBA and Fastidious Anaerobic Agar (FAA) (Media Department, HPA, London) for 24 h and 48 h at 37 °C in anaerobic conditions. In addition, bacterial cultures were grown on Fastidious Anaerobic broth (FAB) for 48 h at 37 °C and re-subcultured on to FAA plates and incubated for 24 h at 37 °C. Target plates were prepared as described above (2.2.2) and the matrix was added prior to analysis. Data acquisition and processing was done as described as above (2.2.3).

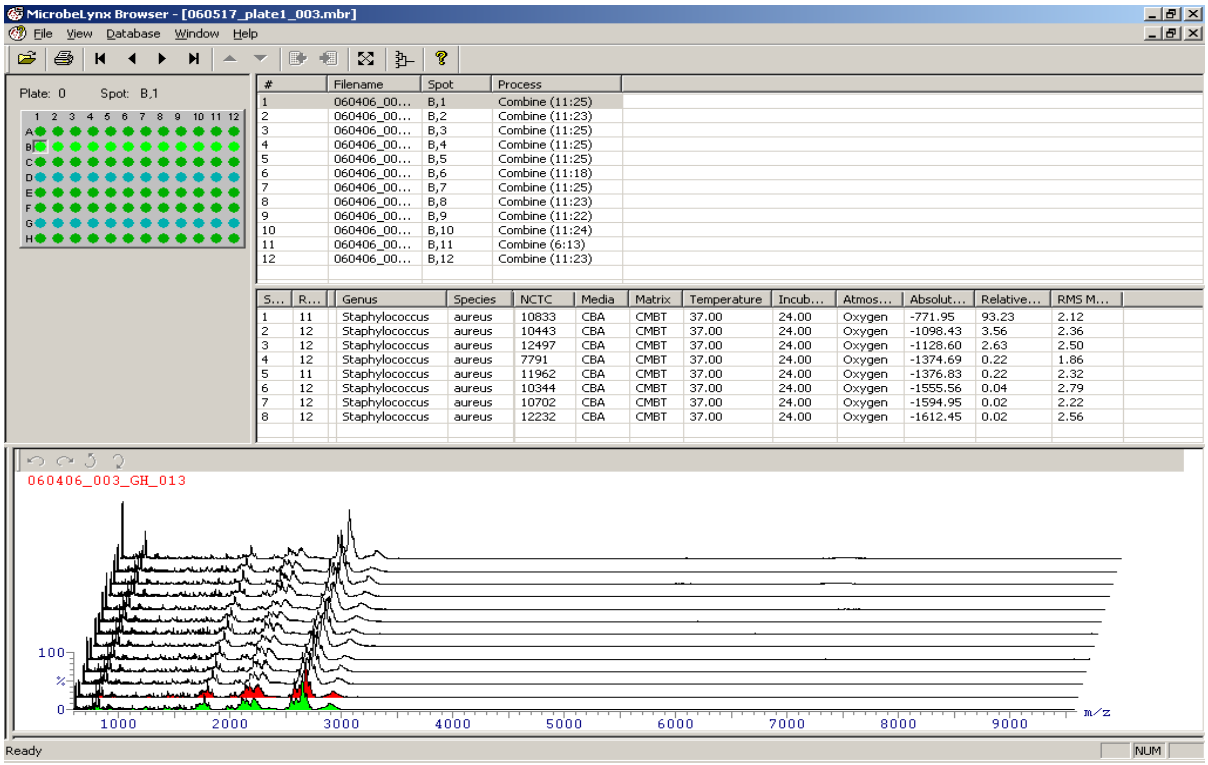


Figure 2.1: MicrobeLynx™ search results window showing the top 8 matches for the unknown isolate against the database. The green spectrum shows the composite profile for all the 12 replicates of the unknown isolate applied to row B. Different colours of the wells on the target plate indicate the probability value: Green- > 80 % relative probability and Blue- up to 80 % relative probability.

2.3 MALDI-TOF-MS analysis of *C. difficile* using the SARAMIS™ database

2.3.1 Optimisation and Sample preparation - *C. difficile*

In order to find the optimal incubation time and a suitable growth medium which gives a high density mass ion spectral profile for *C. difficile*, bacterial cultures were maintained on CBA, Nutrient agar (NA) and Fastidious Anaerobic agar (FAA) plates (Media department, HPA, London) for 24 h and 48 h at 37 °C in anaerobic conditions (AnaeroGen, Oxoid). Forty two isolates (n = 42) on CBA for 24 h, 43 (n = 43) on CBA for 48 h, 10 (n = 10) isolates on FAA for 48 h and 9 (n = 9) isolates on NA for 48 h were subcultured on the first phase.

For the second phase of the study, 22 *C. difficile* isolates (n = 22) on CBA for 24 h, 25 isolates (n = 25) on CBA for 72 h and 16 (n = 16) on CBA for 48 h were analysed including the reference strains of *C. difficile* NCTC 13404, NCTC 13366, NCTC 13307, NCTC 13287 and ATCC 630.

Fresh cells from individual colonies were transferred to a cleaned stainless steel 48 well target plate (Shimadzu Corporation, UK) using a sterile blunt pipette tip in duplicates. Once on the plate, the cells were immediately mixed with 0.5 µl of matrix solution. The matrix solution contained 10 mg/ml of 2, 5-dihydroxybenzoic acid in acetonitrile: ethanol: water (1:1:1) with 0.3 % trifluoro acetic acid (AnagnosTec, Golm, Germany). The plate was left to air dry at room temperature for a few mins. *E. coli* (DSM 1576, CCUG 10979, and ATCC 8739) were used as the positive control and the calibrant (Figure 2.2 and 2.3).

2.3.2 Data acquisition

Mass spectrometric measurements were performed on an AXIMA CFR Plus (Shimadzu Corporation, UK) instrument. Mass spectra were acquired in a positive linear mode with a nitrogen laser of 337 nm. The data acquisition range were from m/z = 2,000 Da to 20,000 Da. All spectra were processed using the accompanying Shimadzu Biotech software and the peak lists were exported to the SARAMIS™ software (AnagnosTec, Golm, Germany). The software was used to match the pattern with comparison function to identify unknown isolates.

2.3.3 Analysis of closely related different *Clostridium spp*. using SARAMIS™ database

Reference and Type isolates closely related to *C. difficile* were obtained from NCTC, HPA, London. These were: *Clostridium difficile* (NCTC 11207), *Clostridium difficile* (NCTC 11209), *Clostridium beijerinckii* (NCTC 13035), *Clostridium chauvoei* (NCTC 8361), *Clostridium bifermentans* (NCTC 6800), *Clostridium tertium* (NCTC 2917), *Clostridium septicum* (NCTC 549), *Clostridium butyricum* (NCTC 6084), *Clostridium paraperfringens* (NCTC 10986), *Clostridium perfringes* (NCTC 13112), *Clostridium histolyticum* (NCTC 7124), *Clostridium tetanomorphum* (NCTC 543), *Clostridium sporogenes* (NCTC 275) and *Clostridium putrificum* (NCTC 4718). The isolates were subcultured on CBA for 24 h at 37 °C.

The target plates were prepared as described above (2.3.1) and MS profiles were analysed by SARAMIS™ software (2.3.2). Further, these isolates were confirmed by 16S ribosomal sequencing to ensure their authenticity (2.2.4).

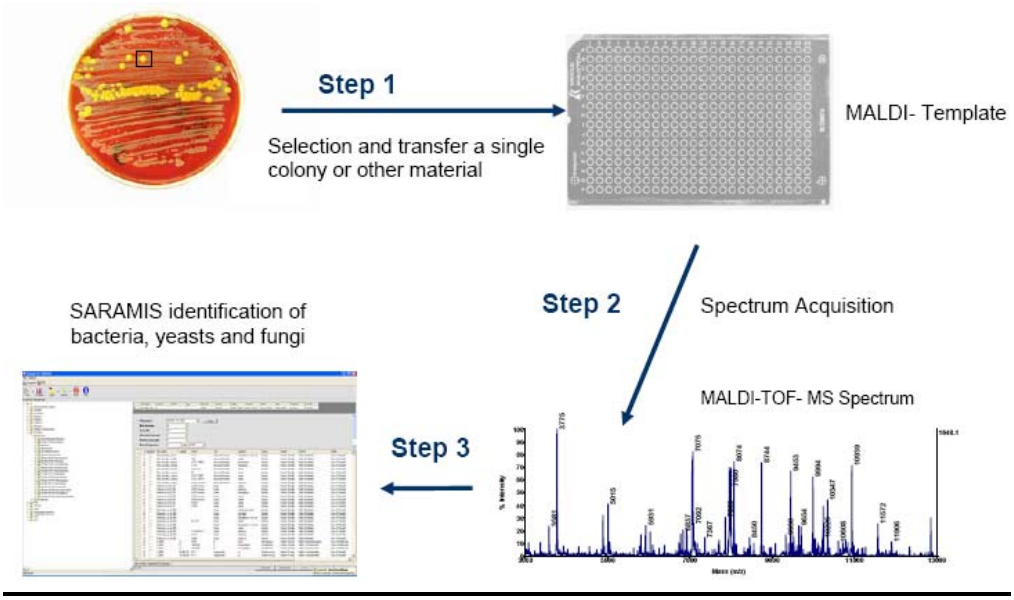


Figure 2.2: Workflow for SARAMIS MALDI-TOF MS Analysis

- 24 h Growth time on CBA
- 48 h Growth time on CBA
- 48 h Growth time on FAA
- 48 h Growth time on NA

Figure 2.3: Sample preparation overview of the different growth times and conditions.

2.3.4 Electron Microscopy of *C. difficile* cells; effect of drying time before MALDI analysis

In order to assess the effect of the matrix on intact cells following addition of the matrix solution, cells taken at different time intervals were processed and observed under electron microscopy.

The electron microscopy was carried out by Dr. Hazel Appleton at HPA, London.

Sample preparation-Batch 1

C. difficile NCTC 13366 was cultured for 16 h on CBA. Cells were collected with a sterile loop and re- suspended in 10 ml of distilled water and the optical density was measured (OD_{600} 0.744). The cell suspension and the matrix solution (AnagnosTec, Germany) were mixed at 1:1, (680µl of the cell suspension and 680 µl of matrix solution) and from this solution 170 µl were mixed with 30 µl of 10 % formaldehyde, to obtain a final concentration of 1.5 % of formaldehyde (Table 2.5). A total of seven samples were prepared at 30S time intervals to cover the drying times used in the standard protocol. Formaldehyde was used as a fixing solution to prevent further reaction of matrix with the cells and prepare grids for Electron Microscopy **(**EM).

Sample No:	Time Intervals (S)	Cell Vol: (µl)	Formaldehyde Vol:(µl) 1.5 %	Total Vol: (µl)
1	0	170	30	200
2	30	170	30	200
3	60	170	30	200
4	90	170	30	200
5	120	170	30	200
6	150	170	30	200
7	180	170	30	200

Table 2.5: The number of samples prepared with different time intervals and the volume of cell suspension and formaldehyde used for each sample.

Sample preparation-Batch 2

CBA plate cultures of 16 h, 24 h and 72 h, *C. difficile* NCTC 13366 isolates were re-suspended in 2 ml of distilled water and optical density measured (OD_{600} 1.022, 1.078 and 1.061 respectively). The cell suspension and the matrix solution were mixed at 1:1, (340µl of the cell suspension and 340 µl of matrix solution) and from this solution, 170 µl were mixed with 30 µl of 10 % formaldehyde, to reach a final concentration of 1.5 % of formaldehyde. A total of four samples per culture were prepared at different time intervals (Table 2.6).

Sample No:	Time Intervals (m)	Cell Vol: (µl)	Formaldehyde Vol:(µl) 1.5 %	Total Vol: (µl)
16 h Culture	0	170	30	200
	5	170	30	200
	10	170	30	200
	15	170	30	200
24 h Culture	0	170	30	200
	3	170	30	200
	6	170	30	200
	10	170	30	200
72 h Culture	0	170	30	200
	3	170	30	200
	6	170	30	200
	10	170	30	200

Table 2.6: The number of samples prepared for different timed interval using cultures, different incubation times, volume of cell suspensions and formaldehyde used for each sample.

Sample preparation-Batch 3

CBA plate cultures of 16 h, 24 h and 48 h *C. difficile* LHI-644 strain was re-suspended in 1 ml of PBS (Phosphate Buffered Saline) (Media department, HPA, London) and the optical density was measured (OD_{600} 2.217, 2.006 and 2.400 respectively). The cell suspension was centrifuged at 21000 × g for 2 min and the supernatant removed. The pellet was re-suspended in 1 ml of the matrix solution. From this solution, 170 µl were mixed with 30 µl of 10 % formaldehyde, to reach a final concentration of 1.5 % of formaldehyde. Five samples per culture were prepared at different time intervals. A control was prepared for all the cultures by suspending the cells in PBS and without the matrix solution (Table 2.7).

Sample No:	Time Intervals (m)	Cell Vol: (µl)	Formaldehyde Vol:(µl) 1.5 %	Total Vol: (µl)
16 h Culture	0	170	30	200
	1	170	30	200
	2	170	30	200
	5	170	30	200
	10	170	30	200
24 h Culture	0	170	30	200
	1	170	30	200
	2	170	30	200
	5	170	30	200
	10	170	30	200
48 h Culture	0	170	30	200
	1	170	30	200
	2	170	30	200
	5	170	30	200
	10	170	30	200

Table 2.7: The number of samples prepared at different timed intervals using cultures at different incubation times, volume of cell suspensions and formaldehyde used for each sample.

2.3.5 Scanning Electron Microscopy of *C. difficile* cells; images before and after MALDI analysis

In order to view the effect of the matrix solution in cell preparations directly on the target plate prior and post mass spectral analysis, 3 target plates were carefully cut into sections, each containing four wells per section. Each section was mounted on to an SEM specimen stub using carbon based electrically conductive double-sided adhesive discs (Agar Scientific). The steel slides were then conductive coated with gold using an Atom Tech Ultra Fine Grain Coating Module (800 series) ion beam coater. The coated slides were examined with an Philips XL30 FEG SEM.

2.4 SELDI-TOF-MS

In the last few years several protocols were developed for protein extractions and analysis of various bacterial pathogens for SELDI-TOF-MS (e.g. Encheva 2005 and Ines 2008). However it is evident that up to the present time universally standardized protocol is not available and prior to any new study, it is necessary to develop an optimum protein extraction procedure. Here several approaches were under taken to obtain a reproducible method for SELDI-TOF-MS analysis of MRSA and MSSA strains of *S. aureus* and to study the microevolution of *C. difficile*.

2.4.1 Different protein extraction methods

1). French Press

The growth from four *S. aureus* plates were suspended in 500 µl of lysis solution and forced through a French Press mini cell (Thermo Fisher, UK) for five times to determine the maximum number of passages required to obtain maximum protein release from cells. During processing, aliquots were taken and protein concentrations estimated using Bradford protein assay (Bradford, 1976). A graph was plotted of concentration against number of passages (Figure 2.4) and demonstrated that the amount of protein extracted reached maximum after three passages and were standardised thereafter.

The same protocol was repeated by re-suspending the cells in distilled water and the protein yield was compared between lysis solution and distilled water.

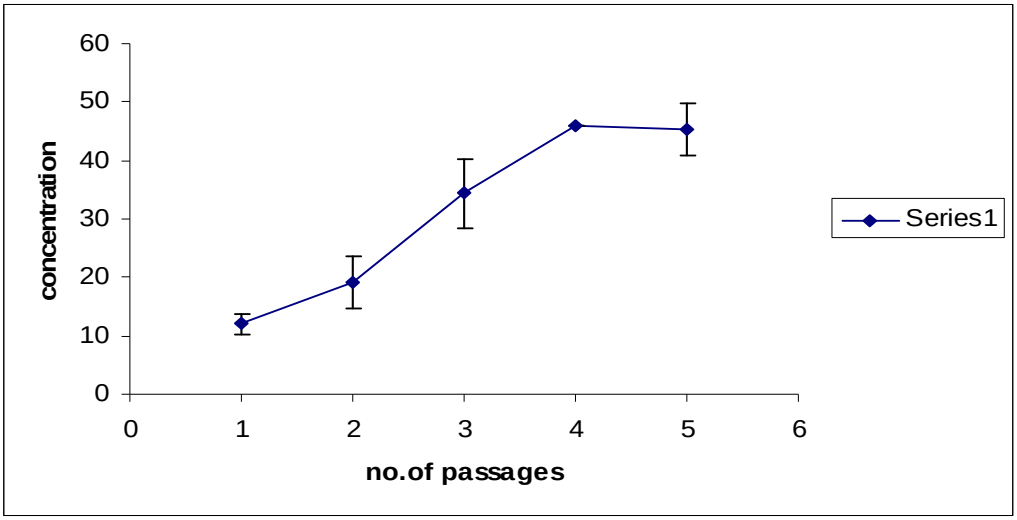


Figure 2.4: The effect of the number of passages through the French Press on protein yield. The results demonstrated that after four passages the protein concentration reaches a maximum.

2) Mickle beating combined with freeze/thawing and different concentrations of Lysostaphin

Growth from four plates of *S. aureus* was harvested and re-suspended in 500 µl of standard lysis solution (8 M urea, 2 % CHAPS (Melfords, UK), 40 mM Tris base, 25 mM PMSF). The samples were tested in duplicate with addition of 15 µl and 30 µl of lysostaphin (Sigma, UK). The cells were mechanically disrupted using the Mickle Cell Disintegrator (Mickle Laboratory Eng.Co Ltd, UK), in the presence of 0.3 g of glass beads, grade 13 (< 105 µm) (Sigma, UK), frozen for 15 min and defrosted for 5 min. This step was repeated twice. The cell debris and the beads were separated from the resulting crude cell extract by centrifugation at 21000 × g for 30 min at 4 °C. The resulting supernatant was collected and the protein content measured using the Bradford protein assay (Bradford, 1976) using serial dilutions of Bovine Serum Albumin (BSA, Sigma UK) as a standard. The same protocol was repeated using distilled water and the protein yield was compared.

2.4.2 Protein extraction for MRSA and MSSA

Protein extraction was carried out as described above by Mickle beating combined with freeze/thawing (2.4.1 (2)) using lysis solution and 30 µl of lysostaphin on each sample.

2.4.3 Different ProteinChip® arrays used for MRSA and MSSA

Three types of ProteinChip® surfaces were tested.

- Weak cationic exchange (CM10)
- Hydrophobic (H50)
- Strong Anionic Exchange (SAX).

2.4.4 Preparation of CM10 array for *S. aureus*

The CM10 array was prewashed twice in 50 % methanol (BDH, Poole, UK), followed by 5 min incubation with 250 µl of 50 mM sodium phosphate (pH = 7) (Sigma, UK) buffer. This step was repeated twice. The buffer was removed from the wells and a total of 150 mg/ml protein of each sample added to separate wells and incubated for 30 min with vigorous shaking. Samples were removed from the wells and 250 µl of 50 mM sodium phosphate buffer (pH = 7) added and left for 5 min and this step was repeated once more. Buffer was removed and each spot was briefly washed with 250 µl of de-ionised water. The array was air dried for a few minutes. One microlitre of the matrix solution containing 14 mg of sinapinic acid (Fluka, UK), 50 % acetonitrile (Sigma, UK) and 0.05 % TFA (BDH, Poole, UK) was added to each spot and analysed using a MALDI-TOF-MS (ProteinChip reader, PBS II, Ciphergen Biosystems). External calibration was performed using the All-in-1 Protein Standard (Ciphergen Biosystems).

2.4.5 Preparation of SAX/Q10 array for *S. aureus*

The SAX/Q10 ProteinChip® array was pre-washed by adding 350 µl binding buffer (50 mM Tris Base pH = 7.8 (Sigma, UK), 2 % Triton X-100 [Sigma, UK]) to each spot and incubating with constant shaking for 15 min. The binding buffer was removed and samples were diluted to 300 mg/ml of protein concentration and 350 µl were added to the wells. The array was incubated for 1 h with constant shaking at room temperature. After incubation, the samples were removed and each spot was washed twice with 350 µl of binding buffer for 5 min. The whole array was washed once in HPLC grade water (Sigma, UK) and left to air dry. One µl of matrix (14 mg of sinapinic acid (Fluka, UK), 50 % acetonitrile (Sigma, UK) and 0.05 % TFA (BDH, Poole, UK)) was added to each spot and analysed using PBS II MALDI-TOF-MS. External calibration was performed using the All-in-1 Protein Standard (Ciphergen Biosystems).

2.4.6 Preparation of H50 array for *S. aureus*

The H50 ProteinChip® array was tested using two different buffers (10 % acetonitrile with 0.1 % TFA (BDH, Poole, UK) and 50 % acetonitrile with 1 % TFA). From each buffer, 250 µl was added to each spot and incubated for 5 min at RT with vigorous shaking. The step was repeated. The binding buffer was removed and 120 µl of sample containing 300 mg/ml proteins were immediately added to each spot. Incubated with vigorous shaking for 30 min. Samples removed from the wells and washed by adding 250 µl of binding buffer for 5 min with agitation. This step was repeated twice. The buffer was removed and washed with 250 µl of HPLC grade water. This step was repeated once. The array was air dried for 30 min and one µl of matrix (14 mg of sinapinic acid (Fluka, UK), 50 % acetonitrile (Sigma, UK) and 0.05 % TFA (BDH, Poole, UK)) was added to each spot and analysed using the PBS II MS (2.4.5). External calibration was performed using the All-in-1 Protein Standard (Ciphergen Biosystems).

2.4.7 Effect of different pHs on CM10 array

Effect of pH for SELDI analysis was tested for CM10 array using two different buffers: 50 mM sodium phosphate (pH = 7 and 8) and 50 mM ammonium acetate (pH = 4 and 6). The ProteinChip® arrays were tested as described above.

2.4.8 Data acquisition parameters

The captured proteins on the surface of the arrays were analysed using a MALDI-TOF- Mass Spectrometer, PBS II (Ciphergen Biosystems. The instrument was equipped with a holder for the ProteinChip® array and used a nitrogen laser (wavelength 337 nm). The instrument was operated in positive ion mode and spectral profiles were collected in the mass range 3000 – 30000 Da. Laser energy used was 200 eV, the plate detector voltage was set at 1900 V, pulse voltage at 3000 V and the source voltage at 20 000 V. Five shots per sample well were acquired and a total of 65 shots were collected from each well. The resulting data were further analysed using Artificial Neural Network (ANNs).

2.5 ANN model parameters

This study utilised a three layer Multi Layer Perception (MLP) ANN model together with the back propagation algorithm. The raw data obtained from the SELDI-TOF-MS consisted of individual *m:z* values with their corresponding relative abundance values between 3,000 Da – 30,000 Da. It is these relative abundance values for each *m: z* values that were used as inputs in the input layer. The network utilised a constrained approach to maximise the efficiency of the analysis whereby two hidden nodes were used in the hidden layer. Two hidden nodes were used in order to amplify the importance of key ions within the mass spectrometry data, while producing accurate predictions and maintaining model generalisation. This approach has been adopted with success on earlier mass spectrometry data (Ball *et al.*, 2002). The output layer consisted of a single node encoded with Boolean representation where MRSA was represented by 1 and MSSA represented by 0.

Prior to analysis and model development, m/z values below 1,000 Da were removed as values below this mass value was deemed to be noisy and unimportant. Also due to the limitations in accurate mass resolution beyond 30,000 Da, everything above this mass value was also removed as reported previously by Ball (Ball *et al.,* 2002). It is also important to select the right combination of learning rate and momentum factor values when creating an ANN model which will control the over training of the model. For this study, a learning rate of 0.1 with a momentum value of 0.5 was used as these values had previously produced encouraging results (Lancashire *et al.,* 2005).

Here, a stepwise approach was taken where the ranking of the important ions were considered. The data set containing 50 isolates of MSSA and 49 MRSA isolates were mixed and fed in to the model which was considered as the inputs. Each ion/data point was passed through 50 sub-models and ions were ranked based on the ability to predict as MRSA or MSSA. After selecting the first representative ion from the first run, the process was repeated by adding other single ions to find the best pair of ions and repeated again until the best sub-set of ions was achieved which had the ability to predict the isolates as MRSA and MSSA. When developing the model, an error threshold value of 0.5 was assigned where the isolates predicted above the value of 0.5 were assigned as MRSA and the rest were as MSSA.

The ANNs analysis was carried out by Dr. Graham Ball at NTU.

2.6 Protein extraction for *C. difficile*

Protein extraction was carried out using the same protocol used for *S. aureus* as described above in section (2.4.1 (2)) using lysis solution. But without lysostaphin.

2.6.1 Effect of lysozyme on SELDI profiles

C. difficile is a gram positive organism that is very resistant to cellular disruption. To aid the process of lysis, cell suspensions were incubated with lysozyme and protein extractions were carried out as described above in section 2.4.1 (2) with the addition of 50 mg/ml of lysozyme per sample.

2.6.2 Different ProteinChip® arrays tested

The ProteinChips®, SAX/Q10, CM10 and H50 were tested using the prepared protein extracts. Based on the preliminary MS analysis, the SAX/Q10 ProteinChip® was chosen for further studies and prepared and analysed as described above (2.4.5).

2.7 SDS-PAGE

2.7.1 Protein extraction methods tested for *S. aureus*

a) Boiling method with sample extraction buffer containing SDS for SDS-PAGE

The growth from four *S. aureus* plates was suspended in sample extraction buffer (0.5 M Tris-HCl, 0.4 % SDS [Sigma UK]) and boiled for 10 min. Samples were left to cool, and centrifuged for 30 min. at 15000 × g. The supernatant was removed and protein concentrations estimated using the by Bradford protein assay.

(b) Protein extraction using lysostaphin and SDS

The growth from five plates of S. *aureus* was collected in 1 ml of deionised water and vortexed for 1 min. Samples were centrifuged for 10 min at 5000 × g and the supernatant removed. A buffer containing 970 µl of 10 mM sodium phosphate (Sigma, UK) and 30 µl of lysostaphin (2 µg/µl) (Sigma, UK) was prepared and 0.5 ml of this solution added to the remaining pellet and incubated at 37 °C for 30 min. Following the incubation 0.5 ml of buffer containing 4 % SDS, 20 % glycerol, 2 % 2- β mercaptoethanol, Tris-HCl pH 6.8 (Sigma, UK) was added and vortexed for 1 min. The samples were incubated in a heating block for 10 min at 100 °C followed by the addition of 0.5 ml of deionised water and incubation for a further 10 min at 100 °C. Samples were left to cool and centrifuged for 30 min at 4 °C, at 15000 × g. Supernatants were collected and protein concentrations estimated using the Bradford assay. SDS-PAGE was carried out on precast NuPAGE® ready-to-run gels (Invitrogen, UK).

2.7.2 *C. difficile*

The same protein samples prepared for SELDI analysis (2.4.2) were used for electrophoresis.

2.7.3 Running the Samples on NuPAGE® ready-to-run gels

The protein samples were denatured in the presence of NuPAGE® LDS sample buffer (106 mM Tris HCl, 141 mM Tris base, 2 % LDS, 10 % glycerol, 0.51 mM EDTA, 0.22 mM SERVA® Blue G250, 0.175 mM Phenol Red pH = 8.5 [Invitrogen UK]) and reduced using 500 mM dithiothreitol (DTT)(Invitrogen UK). The samples were heated in a water bath for 10 min at 70 °C and run on 10 % and 12 % NuPAGE® ready-to-run gels for 40 min with a voltage set at 200 V. A total of 10 µg of the sample was loaded to each well. Rainbow marker and See blue markers were used as the protein standards. NuPAGE® MES SDS (50 mM MES, 50 mM Tris base, 0.1 % SDS, 1 mM EDTA, pH = 7.3 [Invitrogen UK]) was used as running buffer.

2.8 Studies on the intraspecific diversity of *C. difficile* using Variable Number Tandem Repeat (VNTR) analysis

Ninety two isolates including the recently sequenced strain NCTC 13366 (027) strain from Stoke-Mandeville outbreak were used in this part of the study. *C. difficile* isolates were kindly provided by Prof. I. Poxton, MPRL (Microbial Pathogencity Research Laboratory), University of Edinburgh. Another 43 strains were used from a collection from an outbreak in a hospital. This included a recent outbreak strain; ribotype 027 from the Stoke-Mandeville outbreak in 2006 (Table 2.8). The ribotypes of most of the isolates were known prior to VNTR analysis, which was done according to the traditional ribotyping technique (Stubbs *et al.,* 1999). Out of the 92 isolates, 45 belonged to the ribotype 027, 25 were ribotype 001 and the remaining 22 were assigned to different ribotypes.

Strain ID	Source	Ribotype
NCTC 13366	NCTC	Ribotype 027
MPRL 597	MPRL	Ribotype 027
R12087	CD196-Paris	Ribotype 027
R16760	Birmingham	Ribotype 027
R12628	Preston	Ribotype 027
A027	Stoke-Mandeville outbreak	Ribotype 027
G1	Outbreak strain	Ribotype 027
G2	Outbreak strain	Ribotype 027
G3	Outbreak strain	Ribotype 027
G4	Outbreak strain	Ribotype 027
G5	Outbreak strain	Ribotype 027
G6	Outbreak strain	Ribotype 027
G8	Outbreak strain	Ribotype 027
G9	Outbreak strain	Ribotype 027
G10	Outbreak strain	Ribotype 027
G11	Outbreak strain	Ribotype 027
G12	Outbreak strain	Ribotype 027
G13	Outbreak strain	Ribotype 027
G16	Outbreak strain	Ribotype 027
G19	Outbreak strain	Ribotype 027
G20	Outbreak strain	Ribotype 027
G21	Outbreak strain	Ribotype 027
G31	Outbreak strain	Ribotype 027
G32	Outbreak strain	Ribotype 027
G34	Outbreak strain	Ribotype 027
G35	Outbreak strain	Ribotype 027
G37	Outbreak strain	Ribotype 027
G41	Outbreak strain	Ribotype 027
G43	Outbreak strain	Ribotype 027
G45	Outbreak strain	Ribotype 027
G46	Outbreak strain	Ribotype 027
G47	Outbreak strain	Ribotype 027
G49	Outbreak strain	Ribotype 027
G50	Outbreak strain	Ribotype 027
G52	Outbreak strain	Ribotype 027
G54	Outbreak strain	Ribotype 027
G57	Outbreak strain	Ribotype 027
G58	Outbreak strain	Ribotype 027
H 518-G59	Outbreak strain	Ribotype 027
H 521-G62	Outbreak strain	Ribotype 027
H 523-G64	Outbreak strain	Ribotype 027
H 524-G65	Outbreak strain	Ribotype 027
H 526-G67	Outbreak strain	Ribotype 027
H 530-G71	Outbreak strain	Ribotype 027
H 536-G77	Outbreak strain	Ribotype 027
MPRL 3224	MPRL	Ribotype 001
MPRL 3417	MPRL	Ribotype 001
MPRL 3465	MPRL	Ribotype 001
MPRL 3468	MPRL	Ribotype 001
MPRL 3542	MPRL	Ribotype 001
MPRL 3667	MPRL	Ribotype 001
MPRL 3668	MPRL	Ribotype 001

MPRL 3706	MPRL	Ribotype 001
MPRL 3739	MPRL	Ribotype 001
MPRL 3767	MPRL	Ribotype 001
MPRL 3782	MPRL	Ribotype 001
MPRL 4143	MPRL	Ribotype 001
MPRL 684	MPRL	Ribotype 001
MPRL 4147	MPRL	Ribotype 001
MPRL 4178	MPRL	Ribotype 001
MPRL 4340	MPRL	Ribotype 001
MPRL 4342	MPRL	Ribotype 001
MPRL 4702	MPRL	Ribotype 001
MPRL 4773	MPRL	Ribotype 001
MPRL 4778	MPRL	Ribotype 001
G18	Outbreak strain	Ribotype 001
G28	Outbreak strain	Ribotype 001
MPRL 587	MPRL	Ribotype 001
MPRL 549	MPRL	Ribotype 001
MPRL 554	MPRL	Ribotype 001
MPRL 255	MPRL	Ribotype 173
MPRL 585	MPRL	Ribotype 23
MPRL 339	MPRL	Ribotype 23
MPRL 615	MPRL	Ribotype 013
MPRL 845	MPRL	Ribotype 140
MPRL 349	MPRL	Ribotype 125
MPRL 319	MPRL	Ribotype 002
MPRL 557	MPRL	Ribotype 002
MPRL 590	MPRL	Ribotype 165
G23	Outbreak strain	Ribotype 081
MPRL 842	MPRL	Ribotype 005
MPRL 421	MPRL	Ribotype 012
MPRL 688	MPRL	Ribotype 012
MPRL 617	MPRL	Not known
MPRL 602	MPRL	Not known
MPRL 397	MPRL	Not known
MPRL 398	MPRL	Not known
MPRL 841	MPRL	Not known
MPRL 202	MPRL	Not known
MPRL 595	MPRL	Not known
MPRL 381	MPRL	Not known
MPRL 592	MPRL	Not known

Table 2.8: *C. difficile* isolates used for VNTR analysis obtained from MPRL, Edinburgh and outbreak strains from a hospital.

2.8.1 DNA extraction of *C. difficile* isolates using MagNa Pure LC

DNA extraction was carried out on the MagNa Pure LC Robot using the MagNa Pure LC DNA Isolation Kit III (Bacteria, Fungi) (Roche, UK) due to its fast and automated nucleic acid purification. The MagNa Pure LC Robot system with the use of MagNa Pure LC DNA Isolation Kit III (Bacteria, Fungi) helps to isolate high-quality pure DNA. The kit contains three wash buffers which help to remove PCR inhibitors, salts and proteins, a lysis/binding buffer for cell lysis and binding of DNA, proteinase K for protein digestion, magnetic glass particles for binding of DNA and an elution buffer for elution of pure DNA. The bacterial cells were lysed prior to loading the MagNa Pure Robot (Roche, UK) by using a loopful of *C. difficile* cells grown on CBA for 24 h, suspended in 100 µl of PBS (Media Department, HPA, London). To this 130 µl of Bacterial Lysis Buffer and 20 µl of Proteinase K solution were added. The samples were mixed and incubated at 65 °C for 30 minutes followed by incubation at 95 °C for 10 mins to inactivate the cells. Lysates were transferred to the sample cartridge, loaded onto the MagNa Pure machine and DNA was eluted in 100 µl volume of elution buffer according to the manufactures instructions.

2.8.2 VNTR of *C. difficile*: primer design and Tandem Repeat Finder

The Tandem Repeat Finder was used to locate and display tandem repeats in the DNA sequence. Tandem Repeat Database (http://tandem.bu.edu/TRDB.html) is a public repository of information on tandem repeats in genomic DNA and contains a variety of tools for their analysis. In addition, TRDB serves as a centralised research workbench. It provides storage space for results of analysis and permits collaborators to privately share their data analysis (http://cagt.bu.edu/page/TRDB_about). In this study, the genome sequence of strain *C. difficile* NCTC 13307 (630) (GenBank with the accession number AM180355) was used to find the variable regions.

For this study, a total number of 47 loci were tested for tandem repeats, including 10 loci (Table 2.10) which were already published. The remaining 37 loci were identified from *C. difficile* 630 genome. When selecting the loci, a few parameters were considered in order to reduce the complexity. These were copy number, the size of the repeats (pattern size) which

was selected to be 2-45 bp and only loci with percent consensus sequence match of ≥ 95 % were considered for this study (Table 2.9). Each of the selected loci were designated CD1-CD47 with the published loci depicted as CD1-CD10.

	Indices	Pattern Size	Copy Number	% Matches
☑	167124--167172 | [browser]	7	7.000000	100
☐	400403--400585 | [browser]	99	1.800000	97
☑	400436--400528 | [browser]	45	2.100000	95
☑	400905--401171 | [browser]	45	5.900000	96
☑	566910--566950 | [browser]	19	2.200000	97

Table 2.9: Output from the Tandem Repeat finder: 'Indices' represents the position of the repeats on the genome of *C. difficile* 630. The 'Pattern size' is the number of bases present in each repeat while the 'copy number' is number of repeats. The percentage matches indicates the percent sequence similarity between repeats.

2.8.3 Primer designing using Primer3 (http://frodo.wi.mit.edu/)

Primer3 is a web-based computer program used for designing PCR primers required for amplification (Appendix IV). When designing and optimising primers, important factors including melting temperature (T_m), primer length, GC content, 3' stability and likelihood of annealing to or amplifying undesirable sequences were considered. When selecting the primers for this study using Primer3, a product size of up to 700 base pairs was considered in order to yield a high quality PCR product from the ABI sequencer due to the resolution. Melting temperatures between 57 °C – 62 °C were chosen in order to use in multiplex PCR as higher melting temperatures avoids non specific annealing of primers (Figure 2.5).

```
OLIGO            start  len      tm     gc%   any   3' seq
LEFT PRIMER        287   24   61.90   33.33  4.00  2.00
TTCATTTGGAGCTATGGAAATTGA
RIGHT PRIMER       499   21   62.76   52.38  5.00  3.00
CCTTCTCCTGGACTTGCCAAT
SEQUENCE SIZE: 749
INCLUDED REGION SIZE: 749
```

Figure 2.5: Example of output from Primer3 for a specific locus. The left primer consists of 24 bases while right primer contains 21 bases. The total product size including the repeat region will be 749 bases.

2.8.4 Polymerase Chain Reaction (PCR) for VNTR

To increase the throughput of PCR typing and reduce reagent costs, multiple sets of primers can be included in a single reaction tube. Hence, a multiplex PCR was optimised in order to amplify several loci simultaneously (see Appendix V). The forward primers were labelled at the 5' end with a single fluorescent dye [e.g. 5'-carboxyfluorescein (FAM-blue), 4-*N*,*N*-Dimethylaminoethylene amino-*N*-allyl-1,8-naphthalimide (PET-red), 2'–chloro-7'-phenyl-1,4-dichloro-6-carboxyfluorescein (VIC-green) or 2' –chloro-5'-fluoro-7',8'-fused phenyl-1,4-dichloro-6-carboxyfluorescein (NED-yellow)] for all the 47 loci (table 2.6). PCR reactions for all the primer pairs were performed in 12 µl reaction volumes containing 0.5 µl of the diluted DNA (1: 10), 0.03 µl of 1M Betaine (Sigma, UK), 0.5 µl of each of the forward and reverse primers at 5 µM, 4.72 µl of sterile RNAse free water and 6.25 µl of HotStar *Taq* DNA polymerase (Qiagen LTD, UK). The volume of the RNAse free water and the primer concentrations were changed depending on the number of loci amplified in the multiplex PCR (Table 2.10). When the samples were repeated, a single-plex PCR reaction was carried out for each sample.

VNTR analysis was evaluated for its potential to reproducibly amplify, and to precisely size the amplified fragments. The same DNA preparation was used in duplicate VNTR reactions, and analysed on the same gel, and also on different gels. The sizing precision was also tested for single-plex reactions and mulitples reactions. Under all these experimental conditions, the characteristic amplified fragment was reproducibly detected, and all fragment sizes did not vary in any instance by more than ±0.5 bp.

DNA_1:10	0.5	0.5
Master mix-(µl)	**12**	**12**
	Multiplex	**Single-plex**
Distilled water	2.82	4.72
2 X HotStar master mix (1 ×)	6.25	6.25
1M Betaine	0.03	0.03
5 mM Labelled P1_F	0.5	0.5
5 mM-P1_R	0.5	0.5
5 mM Labelled P2_F	0.5	-
5 mM-P2_R	0.5	-
5 mM Labelled P3_F	0.5	-
5 mM-P3_R	0.5	-
Final reaction volume	12.5	-

Table 2.10: Template for preparation of master mix for a single-plex PCR reaction and multiplex PCR reactions. The total reaction volume for each reaction was 12.5 µl.

PCR amplifications were performed on a GeneAMP PCR system 9700 (Applied Biosystems, UK) using the following cycling conditions:

Initial denaturation	94 °C for 15 min – one cycle	
Denaturation	94 °C for 30 sec	35 cycles
Primer annealing	57 °C for 30 sec	
Extension	72 °C for 30 sec	
Final extension	72 °C for 7 min – one cycle	
Hold	4 °C	

2.8.5 Capillary gel electrophoresis on automated DNA sequencer

PCR products were separated by capillary electrophoresis on an ABI 3130xl genetic analyser (Applied Biosystems, UK) using GeneScan™ LIZ600® (Applied Biosystems, UK) as an internal standard for each lane (Figure 2.6). The use of an internal lane size standard enabled automated data analysis and assisted in achieving high run-to-run precision in sizing DNA fragments. The standard was designed for sizing DNA fragments in the 20-600 bp range and provided 36 single-stranded labelled fragments.

For electrophoresis on the genetic analyser, the samples were prepared by adding one µl of the diluted product (1:20 dilution) to 0.4 µl of GeneScan™ LIZ600® size standard and 10 µl of Hi-Di formamide (Applied Biosystems, UK). These were denatured by heating at 95 °C for 5 min followed by cooling to 4 °C. Electrophoresis run conditions were 1.5 kV at 60 °C for 45 min using POP7 polymer as the medium, with a 50 cm capillary array. The GeneMapper software v4.0 (Applied Biosystems, UK) was used to analyse the resulting fragment data.

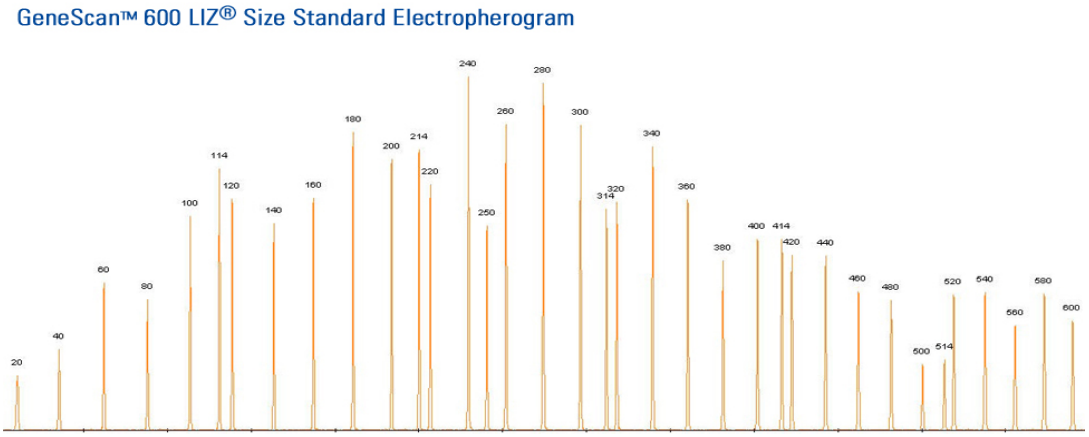


Figure 2.6: Electropherogram of the GeneScan™ LIZ600® size standard (Applied Biosystems, UK).

2.8.6 GeneMapper® software v4.0 analysis

The GeneMapper® Software v4.0 (Applied Biosystems, UK) allows reproducible and highly automated sample analysis, which saves time by facilitating quick data review. Hence, to determine the pattern and size of the fragments obtained from PCR amplification, the Genemapper® software v4.0 was used in this study.

Sample files generated from ABI 3130xl, were transferred to the Genemapper® software as .fsa files for analysis. Analysis parameters that determine the peak detection, sizing and genotyping algorithms were used in the Genemapper® software to analyse the sample files. Sizing of the fragments was done using the GeneScan™ LIZ600® size standard and the 'basic' peak detection algorithm was used. Analysed samples were displayed as electropherograms and visually inspected for the presence and size fragment. Up to eight electropherograms could be viewed at once and the software generates the fragment information for each detected peak in a sizing table with peak size, height, peak area and data point (Figure 2.7). This table can be imported into Microsoft Excel spreadsheet or other software for further analysis.

.

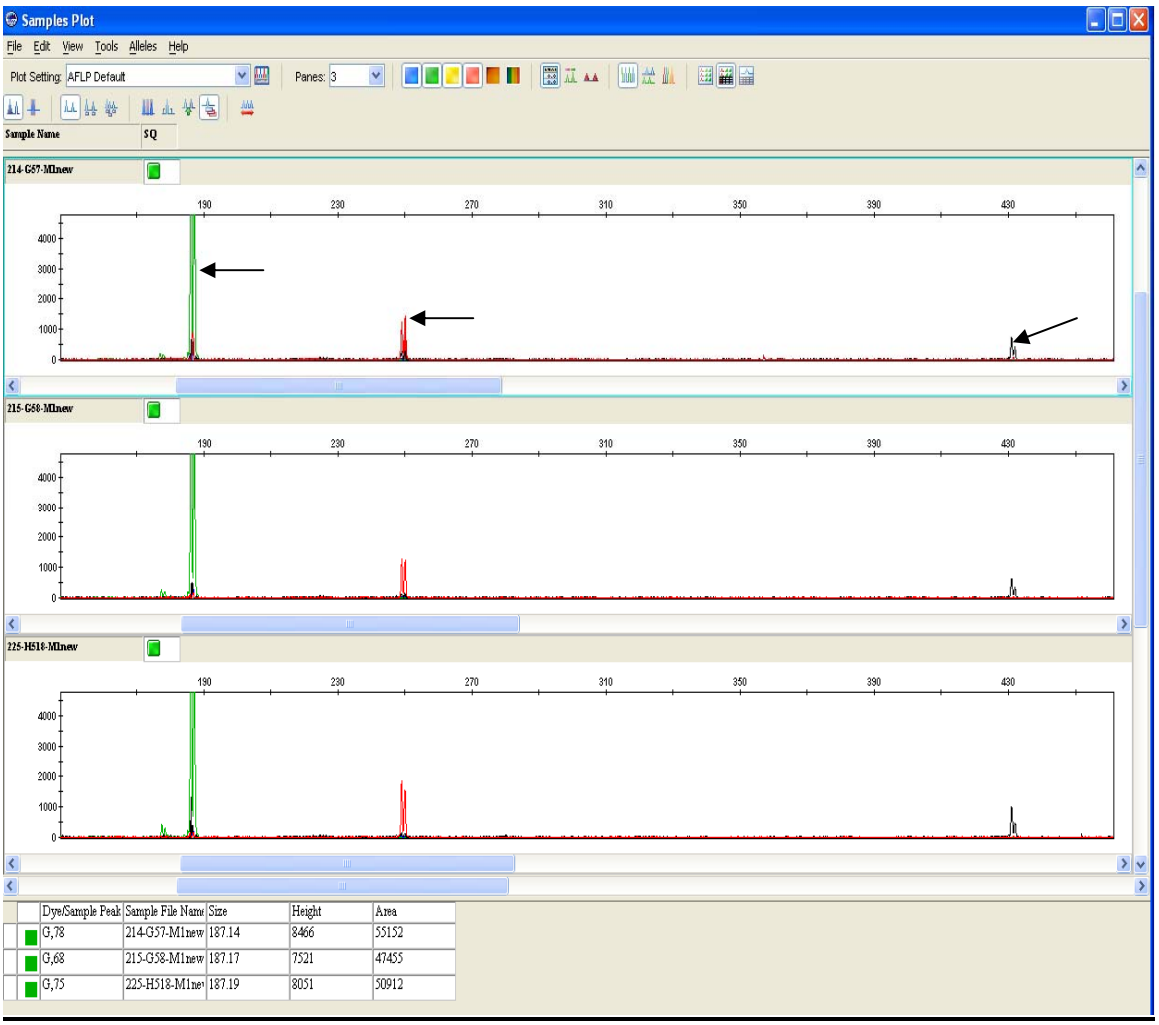


Figure 2.7: Analysis of multiplex samples using Genemapper® software. Analysed samples were displayed as electropherograms and fluorescently labelled amplified fragments could be seen in red, green and black, marked with arrows. Fragment information is displayed at the bottom in a table which could be imported into an Excel spreadsheet.

Locus	Genome position	Repeat sequence	Repeat size
CD1	755721-755950	AAGAGC	6
CD 2	3688632-3688751	ATCTTCT	7
CD 3	3239736-3239835	TATTGC	6
CD 4	167124-167172	ATAGATT	7
CD 5	1954913-1954939	TAT	3
CD 6	664660-664705	TAAAAGAG	8
CD 7	4116072-4116109	TCTTCTTCC	9
CD 8	692929-693015	TATATTGG	8
CD 9	771338-771422	TAAGTATAGAT	11
CD 10	677132-677386	GTAAATAGGATGTAAAA	17
CD 11	2168250-2168268	AAAAAATATA	9
CD 12	804519-804537	CTTCAATAA	9
CD 13	301056-301074	GCT	3
CD 14	788448-788465	TAAATCAGA	9
CD 15	800026-800043	ATAAAGATA	9
CD 16	484804-484820	GAAAAG	6
CD 17	3254341-3254357	TGCTTC	6
CD 18	3014724-3014741	AAACCTTAT	9
CD 19	2167938-2167955	GAACGAATT	9
CD 20	2169070-2169086	AAAGTAT	7
CD 21	2169418-2169435	GATGGCTTA	9
CD 22	804623-804645	TTC	3
CD 23	804652-804699	TTCTTCAGCCTTTTTAGC	18
CD 24	478055-478070	AAAAATG	7
CD 25	543601-543619	TTGCTCATA	9
CD 26	3245854-3245874	TCTTGTATA	9
CD 27	3246180-3246195	AGAATT	6
CD 28	3246752-3246770	TAGATGCAT	9
CD 29	789540-789558	ACTTAA	6
CD 30	794475-794493	ATTAGTG	7
CD 31	3253605-3253625	AATCTTTTA	9
CD 32	252622-252644	AGCAGT	6
CD 33	800719-800741	AT	2
CD 34	799700-799717	TAATAT	6
CD 35	797131-797146	TTCATGA	7
CD 36	881618-881637	AAGAAGAAAA	10
CD 37	881702-881718	GCTATGAA	8
CD 38	883903-883918	AGAAAT	6
CD 39	884179-884194	AAAAAG	6
CD 40	884601-884612	TAG	3
CD 41	886202-886220	AATAAGAGA	9
CD 42	923152-923190	AGTATATTAGTAGTTCTGTA	20
CD 43	623650-623691	TATATGGATAATATCAATTTA	21
CD 44	771250-771319	TAAATATAATCTAA	14
CD 45	3319803-3320164	TTTTATATTAACTATTTTTTT ATTACTTCTATATTATTGTATCA	45
CD 46	3753183-3753574	TTCTTTAGATTAATTTTCTATA CCTAAATTAGTTTATTATAC	45
CD 47	1099662-1099714	AATAAC	6

Table 2.11: All the 47 loci tested for VNTR analysis. CD-1-10 marked in red are already published (van den Berg *et al.,* 2007 and Marsh *et al.,* 2006) and the rest of the loci were identified using Tandem repeat finder and primers were designed using the web-based programme: Primer3.

Chapter 3

Results

3. Results

3.1 MALDI-TOF-MS analysis of *S. aureus*

Spectral profiles of strain HPA 30 grown on different culture media showed striking differences except between CBA and CHOC media. Cells cultured on CBA and CHOC possessed mass ions in the *m*/*z* range of 500-3250 Da (Figure 3.1). Several high intensity peaks were characteristic of the microorganisms grown in CBA in the lower mass range (563, 618, 787, 796 and 825 Da) while a 3012 Da peak had the highest intensity on CHOC grown cells.

Mass ions of cells grown on NA were suppressed especially in the lower mass range and were particularly prominent in the mass range ~ 2000- 3000 Da where only high intensity peaks were 2327, 3009 and 3047 Da. The mass spectral patterns obtained from cells grown on MSA contained the largest number of peaks and spanned the *m*/*z* range 500-3000 Da.

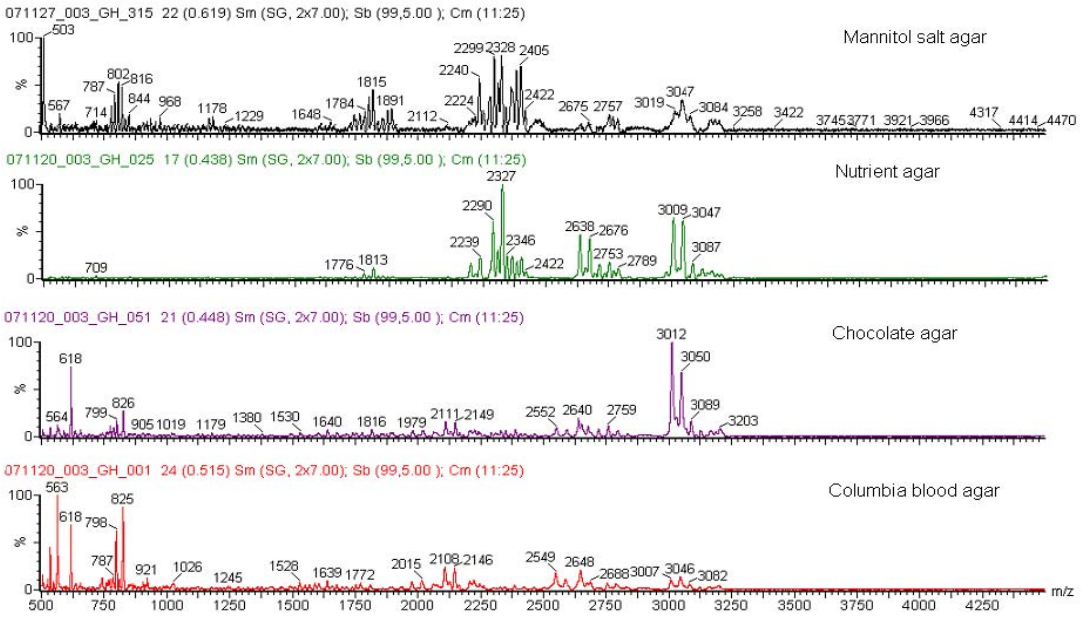


Figure 3.1: Mass spectral profiles of *S. aureus* (clinical isolate HPA 30), when grown on four different types of media viz. mannitol salt agar (MSA), nutrient agar (NA), chocolate agar (CHOC) and Columbia blood agar (CBA). The greatest peak density of mass ions in the m/z 500 to m/z 4000 range was detected when *S. aureus* was grown on MSA and included a number of mass ions with m/z above 3000 Da. CBA and CHOC profiles also contained a broad range of ions in the m/z range of 500–3200 Da. Expanded portions of both spectra (not shown) show consistent spectral features. The least number of m/z peaks were detected from cells grown on NA and concentrated within a narrow section of the spectrum (m/z: between 2239, 2638 and 3047).

Although the MSA gave the largest number of peaks compared to the other growth media, on CBA the mass ion density was still significant. Since the database was developed using CBA medium due to the mass ion density, this was retained as the culture medium for this study.

The MS profiles of strain HPA 30 when grown on CBA showed the largest density of peaks after 24h incubation. With the increasing incubation time, the overall quality of the mass spectral traces deteriorated (Figure 3.2). Therefore, for the rest of the study cultures were grown for 24h at 37 °C.

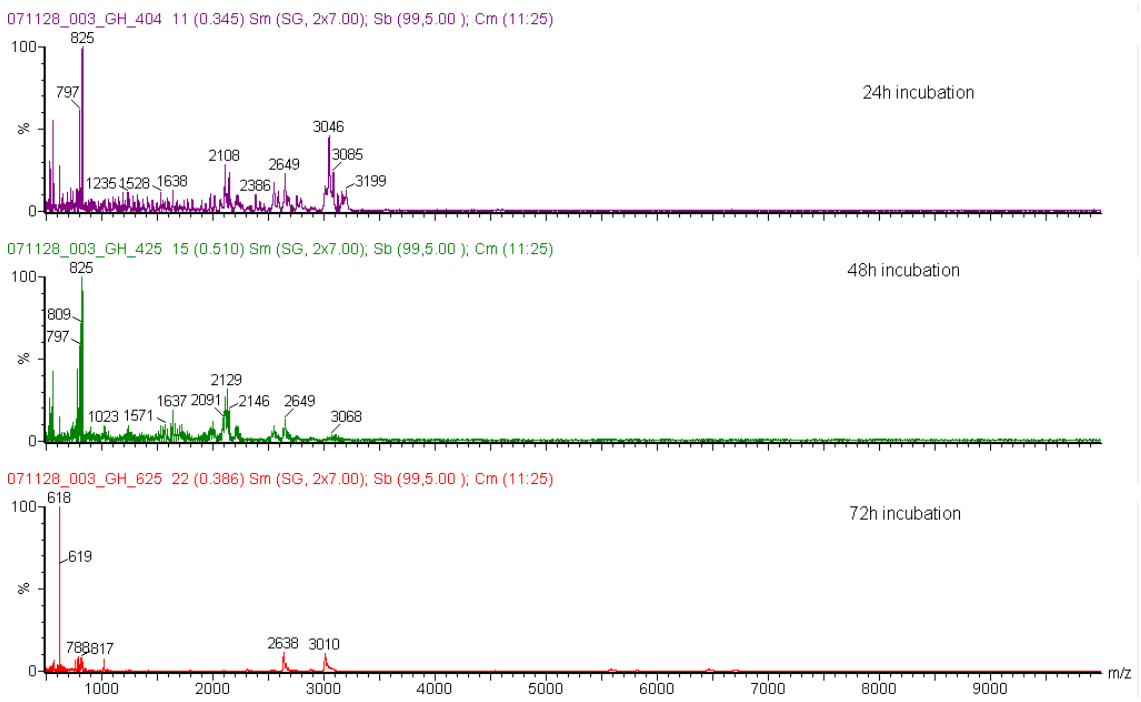


Figure 3.2: Mass spectral profiles of *S. aureus* (clinical isolate HPA 30), obtained after 24, 48 or 72 h of growth on CBA. The greatest number of mass ions was detected on cells grown for 24 h. Consistent mass spectral profiles were obtained after 48 h growth but there was a reduction in mass intensity for some of the characteristic mass ions such as m/z 3046 and 2649. Only a few detectable ions were observed after 72 h emphasising the need to analyse cells within the early to late exponential phase of growth (ca. 24–48 h) to obtain high quality MALDI-TOF-MS profiles.

In the first phase of the study, 95 clinical isolates tested. All except three were correctly identified up to the genus and species levels by MALDI-MS using the 2005 database. The three aberrant isolates (HPA 80, HPA 547 and HPA 549) were identified as *Streptococcus pyogenes, S. haemolyticus* and *S. epidermidis* respectively (Figure 3.3).

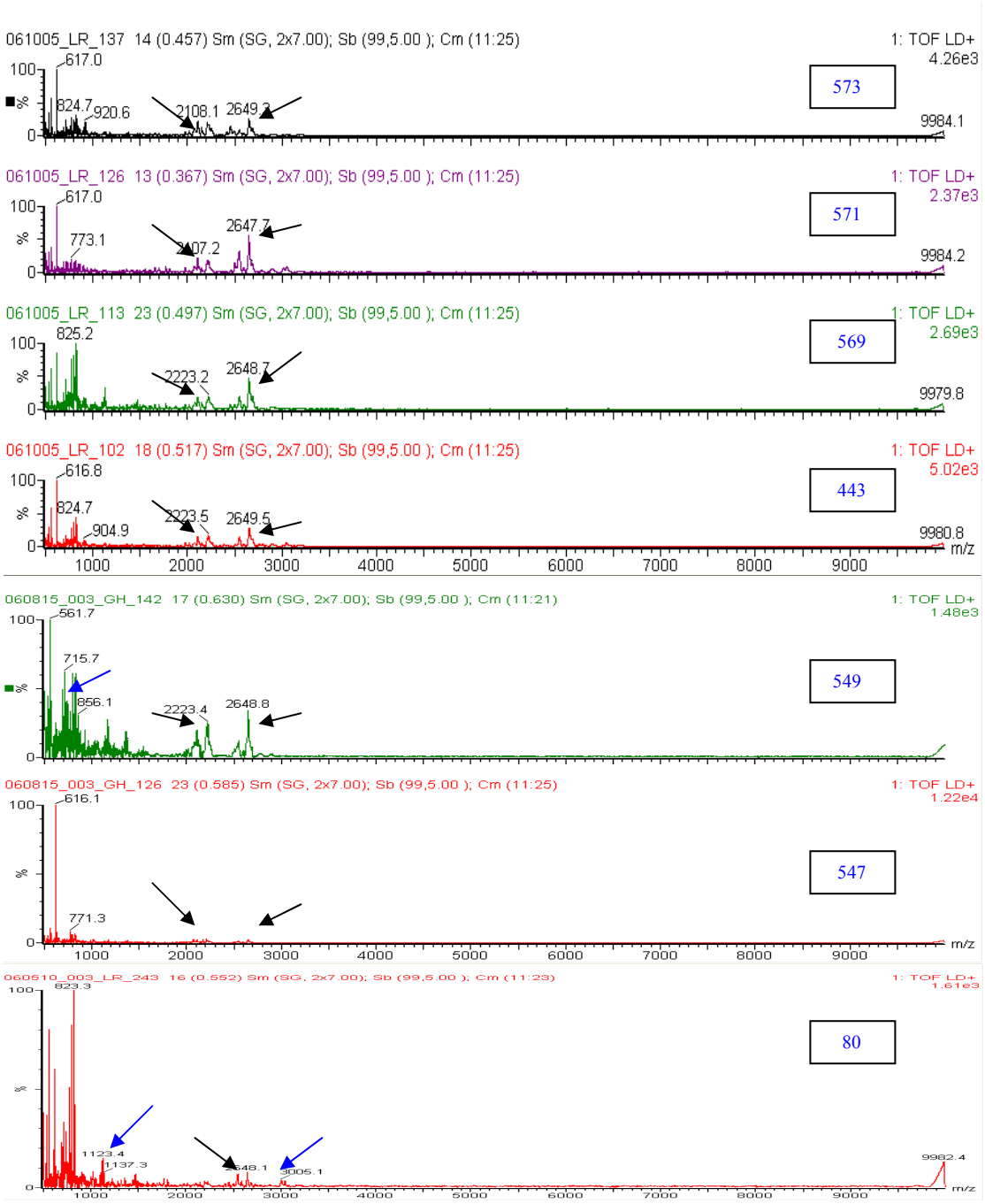


Figure 3.3: Comparison of the profiles of four clinical isolates (573, 571, 569 and 443) which were correctly identified as *S. aureus* and isolates (549, 547 and 80) which was incorrectly identified as *S. epidermidis, S. haemolyticus and S. pyogenus.* All the profiles were visibly similar, shared common peaks except the three isolates, which were incorrectly identified by the software (blue arrows).

To validate the methodology and assess the inter-laboratory reproducibility, a parallel study was performed in collaboration with Manchester Metropolitan University (MMU). All isolates were re-analysed on a different instrument using the same protocol except the data were searched on MicrobeLynx^{TM} 2006 database which contained a further 700 isolates.

Four of the isolates did not correlate and were different to that reported using the 2005 database. These isolates were contaminated upon arrival in Manchester (Table 3.1). However, when incorrectly identified isolates at MMU were re-analysed using the MALDI instrument at the HPA, they were correctly identified as *S. aureus.* All the isolates incorrectly identified using the MALDI-TOF-MS 2005 database were further characterised using sequence analysis of the 16S ribosomal gene. The latter confirmed their identity as *S. aureus* and indicated that the MALDI-TOF-MS 2005 database matches for these strains were incorrect (Appendix I).

ID	**MALDI-ID (HPA)**	**MMU-ID (1st sub culture)**	**MMU-ID (2nd sub culture)**
HPA 248	*S. aureus*	*Proteus mirabilis*	*Proteus mirabilis*
HPA 260	*S. aureus*	*S. warneri*	*S. warneri*
HPA 299	*S. aureus*	Mixed Results	Mixed Results
HPA 356	*S. aureus*	Mixed Results	*S. aureus*
HPA 410	*S. aureus*	Mixed Results	*S. aureus*
HPA 563	*S. aureus*	*Micrococcus luteus*	*S. aureus*
HPA 569	*S. aureus*	Mixed results	*Pseudomonas aeruginosa*

Table 3.1: Comparison of the search results of clinical isolates which matched incorrectly with the 2006 database at Manchester Metropolitan University (MMU) after two subcultures. Results marked in red indicate the samples may have been contaminated during transfer while the blue colour indicates an incorrect match even after two sub cultures.

In the second phase of the study, 35 isolates out of the 39 provided by the Staphylococcal Reference Unit were correctly identified to the species level at the first on the MALDI (Appendix II). The remaining four isolates were incorrectly identified as *S. epidermidis* but were correctly identified as *S. aureus* after the re-analysis by MALDI-MS (Table 3.2).

No. Isolates	Staphylococcal Reference Unit identification	2006 MALDI Database identification	2nd MALDI Run for incorrect samples
35	*S. aureus*	*S. aureus*	-
4	*S. aureus*	*S. epidermidis*	*S. aureus*

Table 3.2: Key features of the clinical isolates obtained from the Staphylococcal Reference Unit, HPA, London and their corresponding identifications obtained by MALDI analysis. A total of 39 isolates were analysed using 2006 MS database. Thirty five isolates were correctly identified at the first run while four misidentified as *S. epidermidis* using the MicrobeLynx™ software. These four isolates were re-run and correctly assigned to the species *S. aureus*.

Comparison of the spectral profiles of the clinical isolates against the type isolates, distinct peaks could be seen in the mass range ~ 800 - 3500 Da, which are characteristic of the genus *Staphylococcus* which is similar to what was reported by Edward-Jones *et al.* (2000), Du *et al.* (2002) and Bernardo *et al.* (2002). Few of the clinical isolates had additional peaks (~ 3200) which maybe specific to those isolates that have m/z values in the mass range of ~ 3000- 4000 Da. The three isolates that showed lower probability scores and mixed matches against the database search, shared similar profiles to *S. aureus* but with less intense peaks in the same mass range and possessed additional mass ions (Figure 3.4).

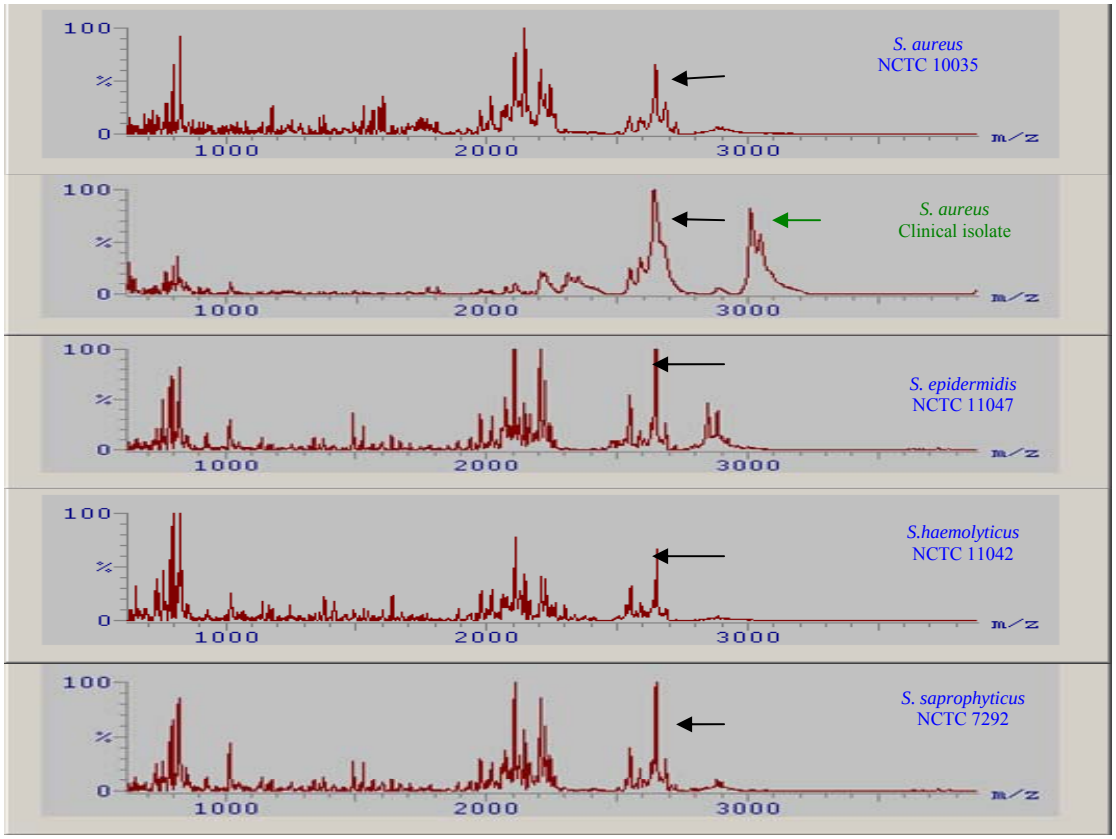


Figure 3.4: Comparison of the MS profiles of clinical isolates against the database of Reference/Type strains. Characteristic peaks were present in all the isolates within the mass range of 500-3500 Da, which are specific to the genus *Staphylococcus.* Common peaks were observed in the m/z range of 500 - 1000 Da and 2000 – 3000 Da for both the type strain of *S. aureus* and clinical isolates. Black arrows indicate the peaks which are specific to the genus *Staphylococcus* and the green arrow shows additional peaks in MS spectrum of a clinical isolate.

3.2 MALDI-TOF-MS analysis of *C. difficile*

Isolates that date back to 1970s and more recent outbreak isolates were analysed by MALDI-TOF-MS. Among these were three representative isolates designated B1 (1979), T (1985) and L (2006 = ribotype 027). A database search of the spectra was performed against the MicrobeLynx^{TM} 2006 version of the database.

Using this database spectral profiles obtained from different culture media and different incubation times were unable to delineate *C. difficile*. However, when the growth medium was changed from Colombia blood agar (standard media used to compile database) to Fastidious Anaerobe Agar (FAA), identification reached to genus level in a few cases (Table 3.3). Although the database predicted results were mixed against the three isolates, they showed similar profiles and similar mass ions/ peaks (~ 1400 and 2600 Da) when compared to the reference strain of *C. difficile*, suggesting a close relationship. However, because of the poor resolution of this species mix up using this protocol (CMBT) (Figure 3.5), in this following section a new approach was applied (SARAMIS™).

Strain Name	Growth Medium	Incubation Time	Predicted ID
B1	CBA	48 hr	*Actinomyces howellii*
L	CBA	48 hr	*Rhodococcus erythropolis*
T	CBA	48 hr	*Sarcina ventriculi*
B1	FAA	48 hr	*Clostridium scatologenes*
L	FAA	48 hr	*Clostridium scatologenes*
B1	CBA	24 hr	*Lactobacillus rhamnosus*
L	CBA	24 hr	*Sarcina ventriculi*
T	CBA	24 hr	*Rhodococcus erythropolis*
B1	FAA	24 hr	*Clostridium scatologenes*
L	FAA	24 hr	*Clostridium scatologenes*
B1	FAB/ FAA	48 hr/ 24 hr	*Clostridium scatologenes*
L	FAB/ FAA	48 hr/ 24 hr	*Clostridium scatologenes*

Table 3.3: MALDI-TOF-MS predicted results for *C. difficile* using CBA, FAA and FAB as growth media and different incubation times.

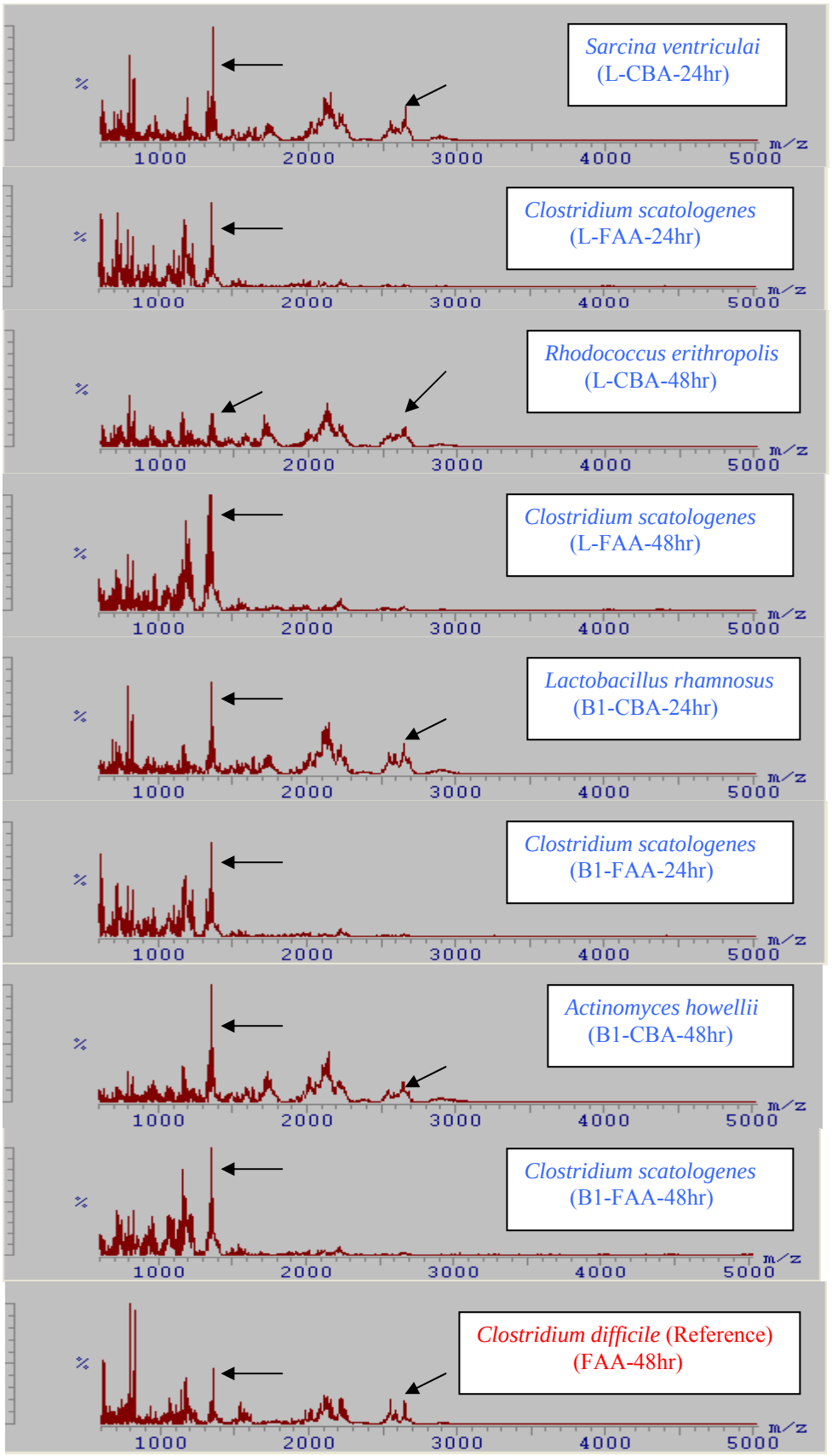


Figure 3.5: Mass spectral profiles of *C. difficile* strains L and B1 against an authentic reference strain. These strains were cultured on CBA and FAA media. The profiles show a large number of common mass ions (black arrows). But the MirobeLynx database matches do not correspond.

3.3 *C. difficile* analysis using SARAMIS™ database

In attempt to improve the level of differentiation of this taxon, a new protocol involving the use of DHB, TFA and ACN (see methods 2.3.1) was applied.

3.3.1 Preliminary results using SARAMIS™ database

In a preliminary study, cells were cultured in three different media; CBA as before and in addition on NA and FAA between 24 - 48 hrs. A striking improvement was obtained both in quality of the MS profiles and the reproducibility. Thus in spite of the variation in the composition of the medium, there was a high degree of compatibility between the spectra was observed (Figure 3.6).

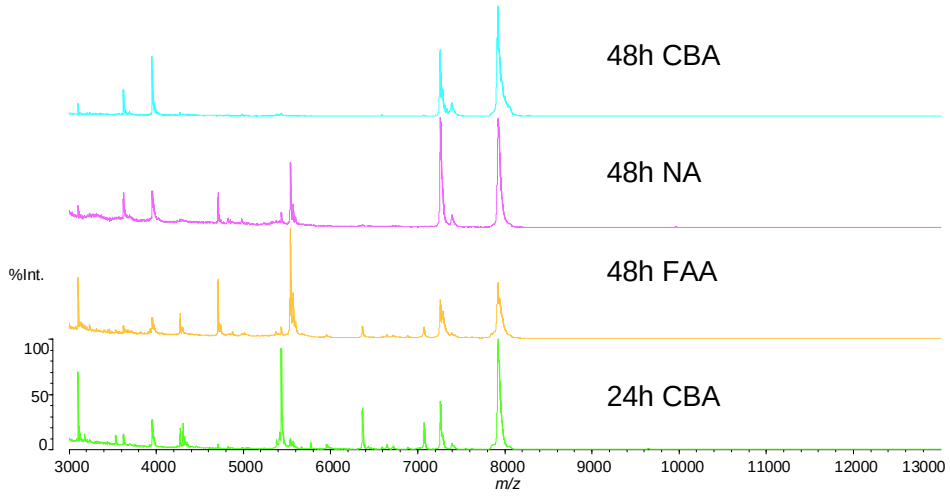
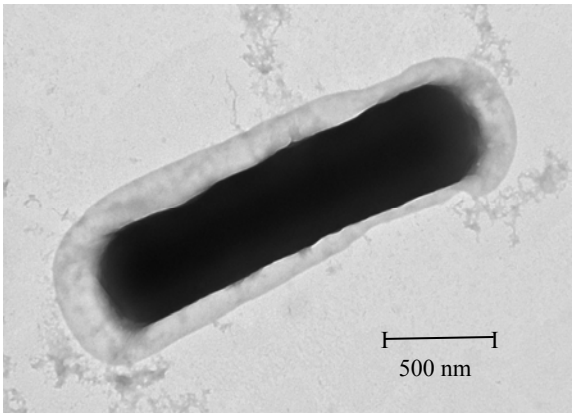


Figure 3.6: Mass spectra of *C. difficile* (MPRL 1037) using the SARAMIS method. Spectra from different culture media showed similar profiles even though the cultures were grown on different media (CBA, FAA and NA) and different incubation times (24 h and 48 h).

3.3.2 Electron Microscopic analysis of *C. difficile*; effect of drying time of the matrix on cells

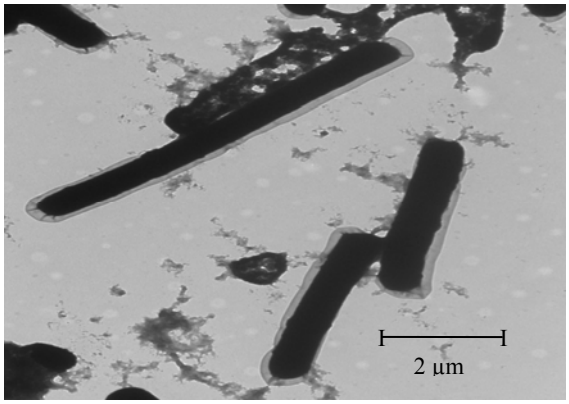
The previous Waters® system used two different matrices CMBT and CHCA for both gram positive and negative bacterial analysis respectively. However with the new SARAMIS protocol, improvement in data using the 'new' matrix solution, (2, 5-dihydroxybenzoic acid in acetonitrile: ethanol: water (1:1:1) with 0.3 % trifluoro acetic acid) was observed. Hence, the interaction of DHB with the cell was examined using transmission and scanning electron microscopy (EM). The following were images obtained from three sample preparations at different timed intervals (Figure 3.7 and 3.8). The samples prepared using PBS gave clear images compared to the samples prepared with water (Figure 3.7 c).

In the controls of 16 h, 24 h and 48 h cultures prepared using PBS, before the addition of matrix, clear intact viable cells could be seen with its capsule around the cell wall and flagella (Figure 3.7 a, b and c). At 48 h culture samples revealed a mixture of intact viable cells and non viable cells (Figure 3.8 c). The results of the cells' reaction to the matrix, with different drying times revealed that the drying time had no major effect on the cell as they remained intact even after five minutes of the addition of the matrix solution. The only effect observed was the disruption of the capsule around the cells (Figure 3.9). This maybe due to the higher polarity of the solvents used in the matrix solution i.e. ACN, ethanol and TFA. In Waters® system, the matrix solutions used were only able to ionise the surface molecules and detected peaks were mainly from the surface molecules of the intact cells (Keys *et al.,* 2004). But this result indicates that the 'new' matrix, DHB is able to disrupt the surface layer i.e. capsule and the observed peaks in the MS are mainly from the intracellular proteins.



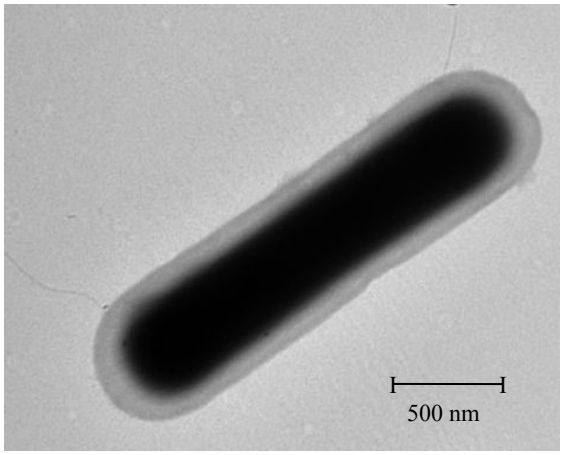
Water

(a)



Water

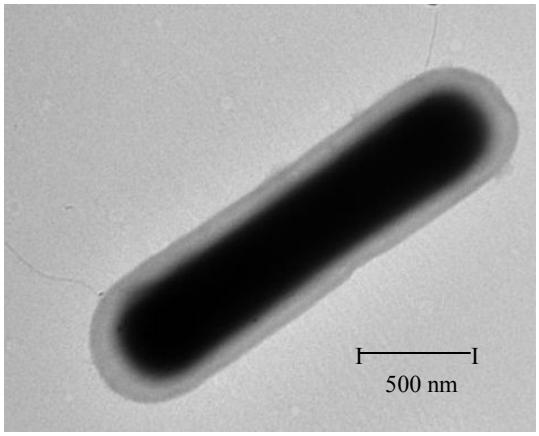
(b)



PBS

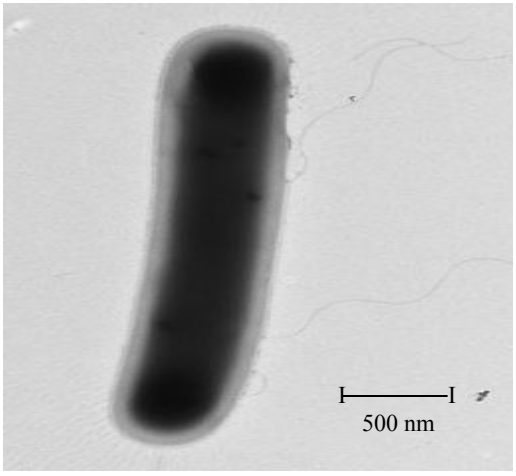
(c)

Figure 3.7: The 16 h untreated samples of *C. difficile* (without matrix) for the three sample preparations (a, b and c). Cell suspension prepared using PBS (c) showed a clear image of the intact cells.



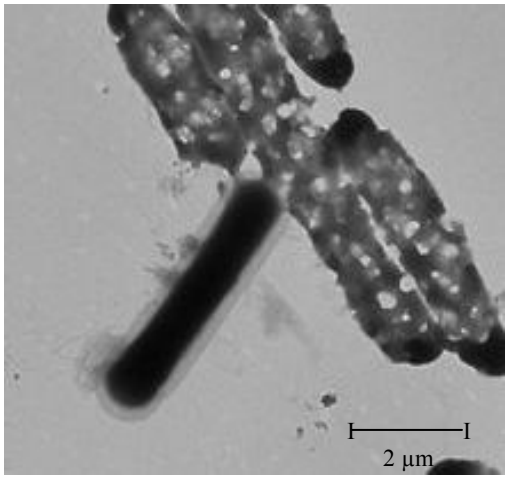
16 h

(a)



24 h

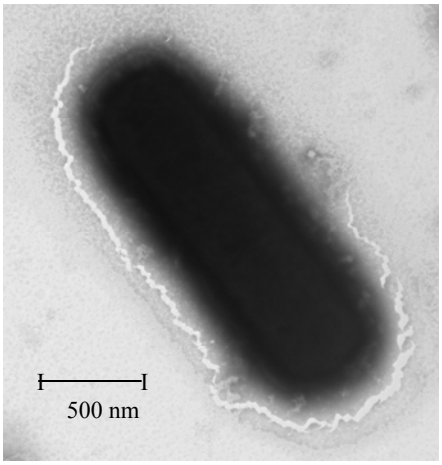
(b)



48 h

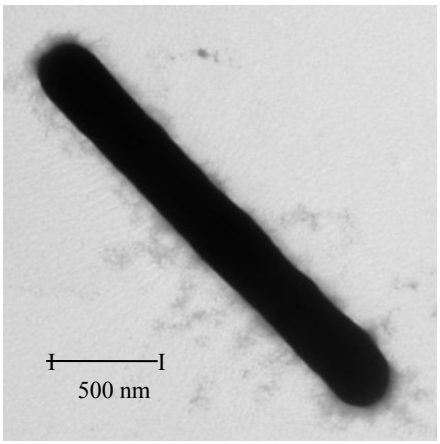
(c)

Figure 3.8: The *C. difficile* untreated samples of 16 h, 24 h and 48 h cultures prepared using PBS (a, b, and c) respectively . Clear intact cells could be seen for all three incubation times before adding the matrix. However the 48 h culture sample revealed a mixture of non viable and viable cells.



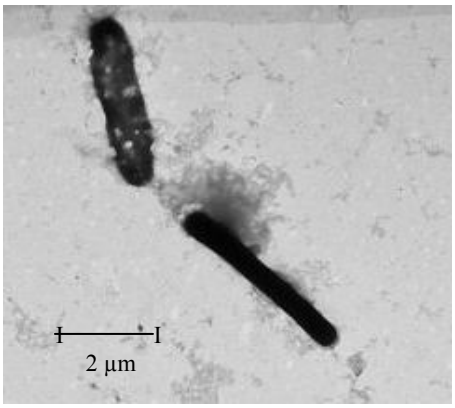
16 h – 150s

(a)



24 h – Time 0

(b)



48 h – 5 min

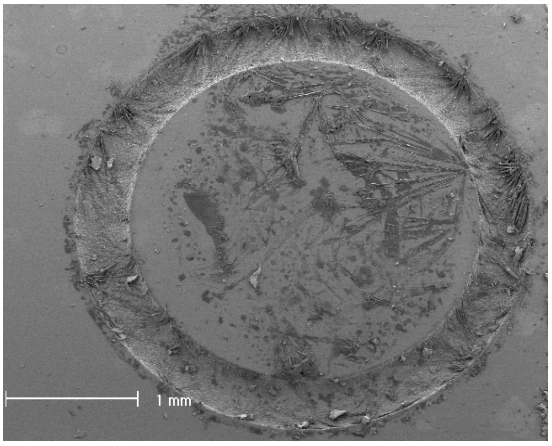
(c)

Figure 3.9: Electron micrographs of different sample preparations of *C. difficile* after adding the matrix: (a) 16 h cells after 150s of matrix addition, (b) 24 h cells prepared by PBS at time 0 and (c) 48 h cells prepared by PBS after five min. of matrix addition. Cells remain intact even after the addition of matrix although damage to the outer polymeric layers was evident.

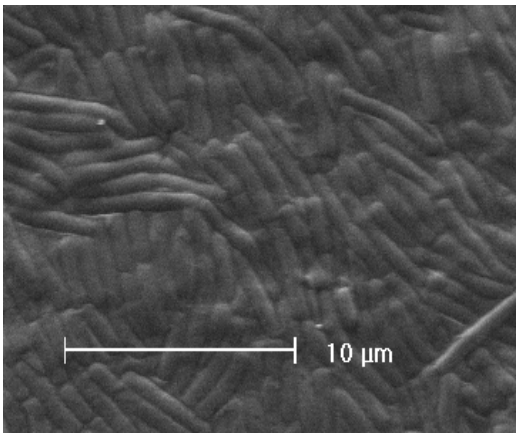
3.3.3 Scanning Electron Microscopy of *C. difficile* on MALDI target plates: before and after MALDI analysis

The scanning electron micrographs were taken on the MALDI target plates, before and after the MALDI analysis, in order to observe the cells' reaction to the matrix solution as well as the MALDI analysis. Before the MALDI analysis and after the addition of the matrix solution, the intact cells were observed. However, instead of a uniform layer of cells across the target well, cells clumped, so that areas of well, showed no cells (Figure 3.10 a). Closer inspection of these densely placed areas (Figure 3.10 b) showed cells that were mostly in parallel. Following the MALDI analysis, the cells appeared less uniform and showed a metallic sheen indicative of disruption of the surface polymers (Figure 3.10 c).

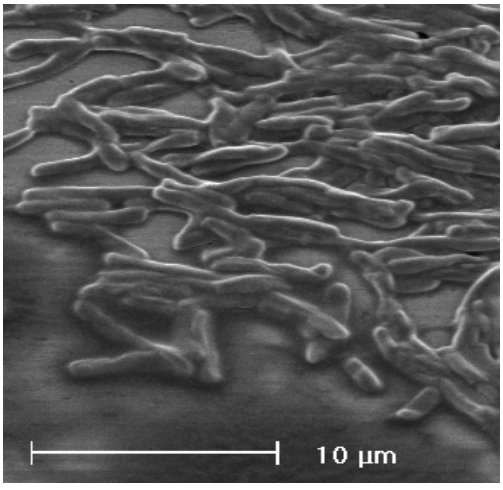
This highly polar solvent (CAN and TFA) would be expected to completely disrupt the cells, but it may have rapidly crystallised with evaporation and held the cells in a 'frozen' state. Consequently, mass spectra of cells were remarkably stable and revealed many intracellular proteins such as ribosomal proteins.



(a)



(b)



(c)

Figure 3.10: Scanning electron micrographs of cells on a MALDI target plate well: (a) A single well after MALDI analysis, (b) Target plate well with *C. difficile* culture on it before the analysis and (c) *C. difficile* cells on the target plate after MALDI analysis.

3.4 Mass spectral analysis of *C. difficile* isolates using SARAMIS™ database

In the first phase of this part of the study, 54 isolates were cultured under anaerobic conditions on CBA for 24 h and analysed by MALDI-MS and database searches performed using SARAMIS™. Except for two isolates, the rest were correctly identified as *C. difficile* (Figure 3.11 and 3.12). The two isolates incorrectly identified as *S. warnerai* and *Propionibacterium acnes* were due to the presence of mixed cultures (Figure 3.14). Compared to the isolates grown on FAA and NA at 24 h and 48 h, the isolates grown on CBA for 24 h produced better mass spectra.

In the second phase of the study, a total of 63 isolates including the reference strains were analysed by MALDI-MS in duplicates. All the isolates were correctly identified as *C. difficile* and a cluster analysis was performed against the available *C. difficile* data in the database. All isolates showed high affinity and were conserved in a single cluster.

The isolates correctly identified by the software were given as an automatic confidence level in percentage. For *C. difficile*, most of the isolates were identified with a confidence level of 80 - 99 % when analysed in duplicates (Appendix III) where most of the significant peaks were seen at the mass range of 2000- 8000 Da (Figure 3.11 and 3.12).

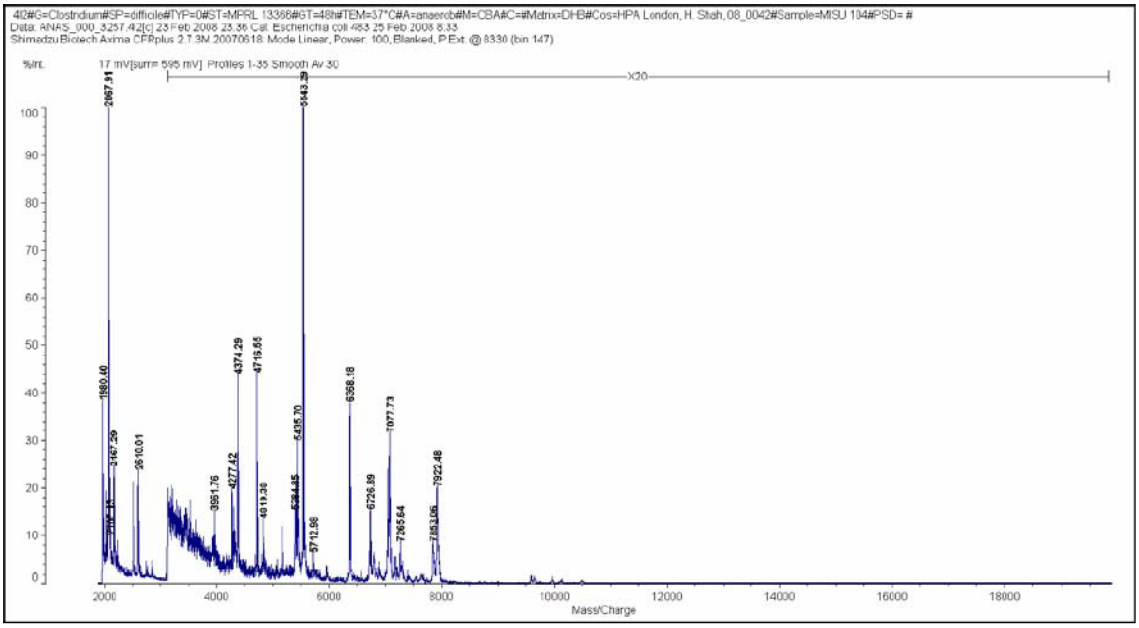


Figure 3.11: Example of a correctly identified isolate (MISU 104) as *C. difficile* grown on CBA for 24 h by the SARAMIS™ database.

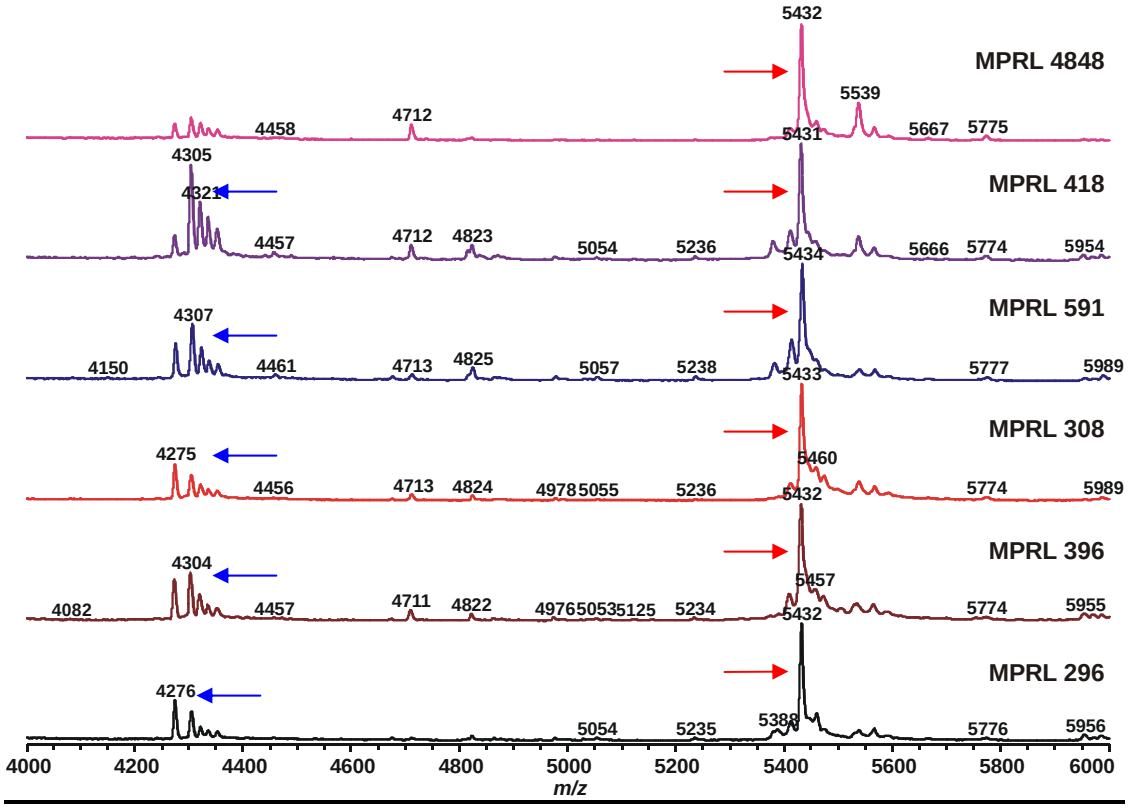


Figure 3.12: Mass spectra of different *C. difficile* isolates. Significant speaks could be seen for all *C. difficile* isolates at ~ 4200 - 4300 Da and at ~ 5400 Da (red and blue arrows).

Further analysis of the spectra of *C. difficile* isolates revealed characteristic peaks among isolates at ~ 4200 Da, ~ 4300 Da, ~ 5400 Da (Figure 3.12) and ~ 6300 Da, ~ 6640 Da and ~ 6720 (Figure 3.13). Also the presence of similar peaks with a single mass shift could be observed, which may be due to the resolution of the instrument (Figure 3.13).

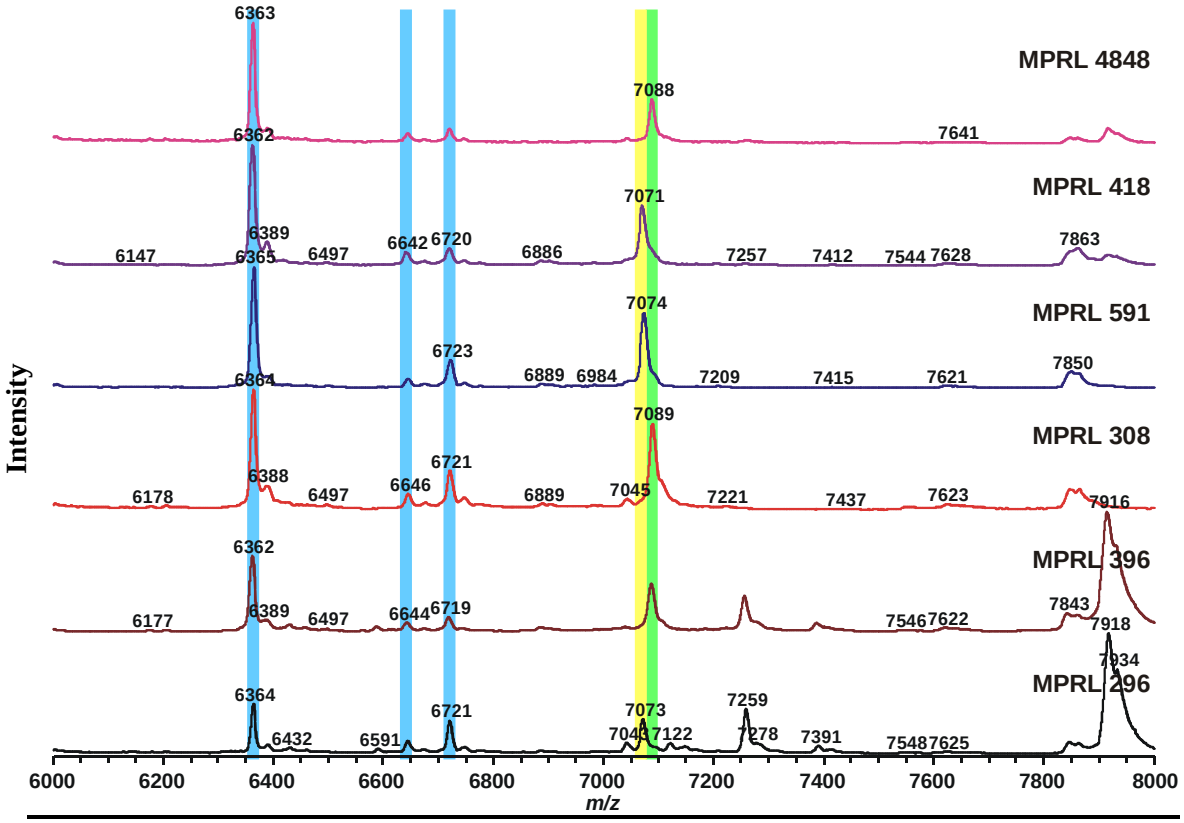
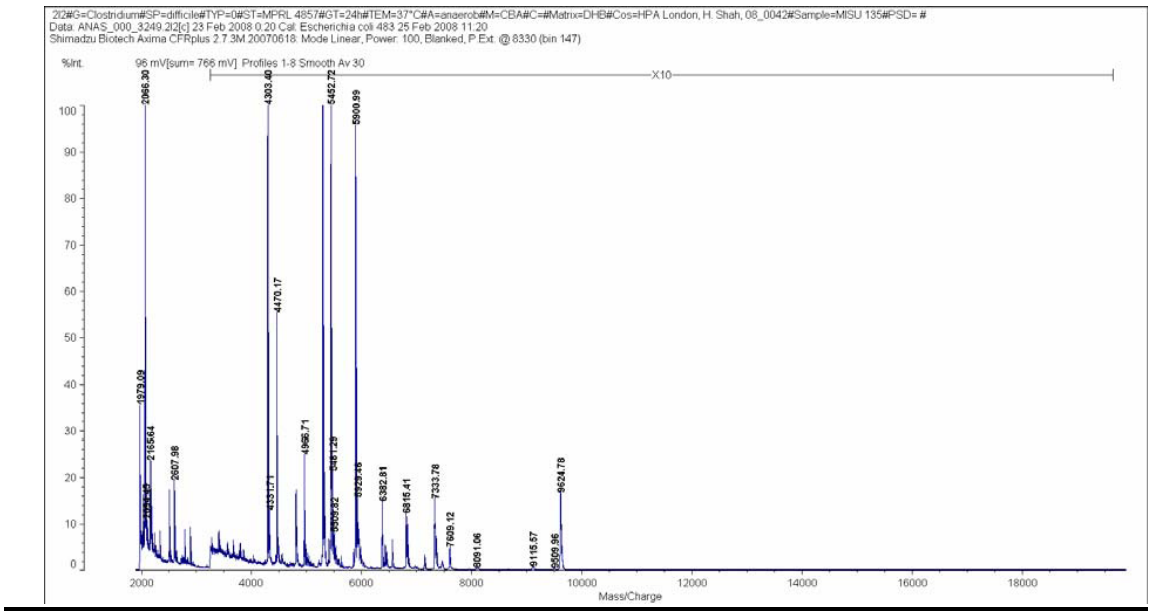
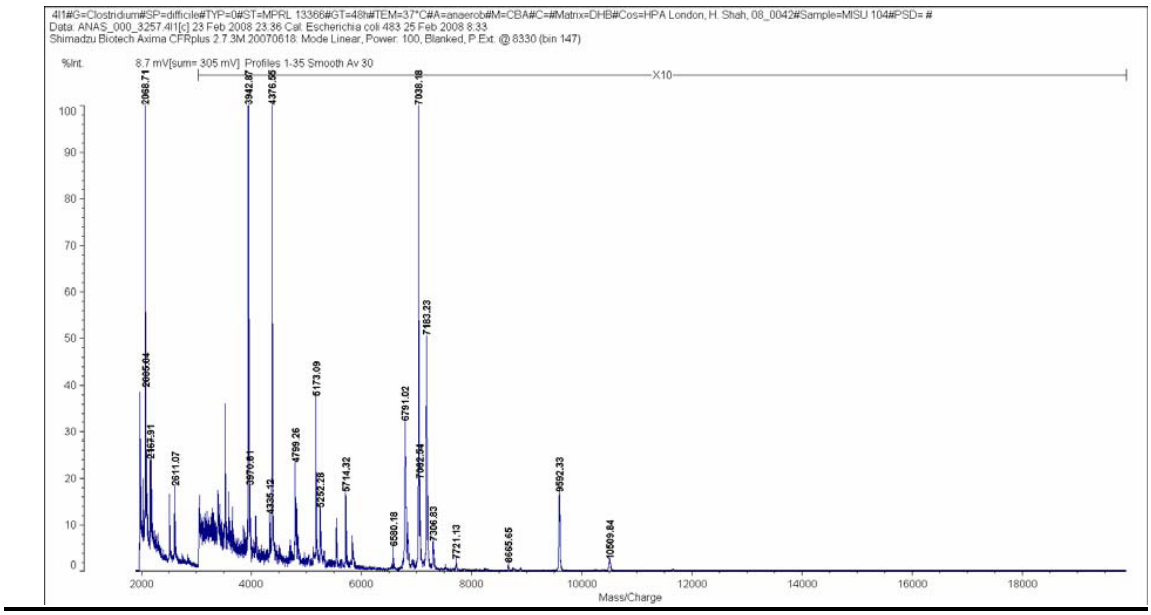


Figure 3.13: Mass spectral profiles of *C. difficile* isolates: Blue lines indicate the peaks which are common to all the isolates. The green and yellow lines indicate peaks that are displaced perhaps due to the resolution of the instrument.



(a)



(b)

Figure 3.14: MALDI-TOF-MS, SARAMIS™ mass spectral profiles of incorrectly identified isolates: MISU 135 (a) and MISU 104 (b) as *S. warnerai* and *P. acnes* respectively.

3.4.1 Differentiation of *C. difficile* from other closely related Clostridial species

Fourteen Clostridial species that are closely related to *C. difficile* were analysed using the SARAMIS™ database. Before MALDI analysis, the identity of species was verified by 16S rRNA sequencing. All the fourteen species were correctly identified by the SARAMIS™ software with a confidence value of more than 65 % (Figure 3.15). Few common peaks could be observed among the different species.

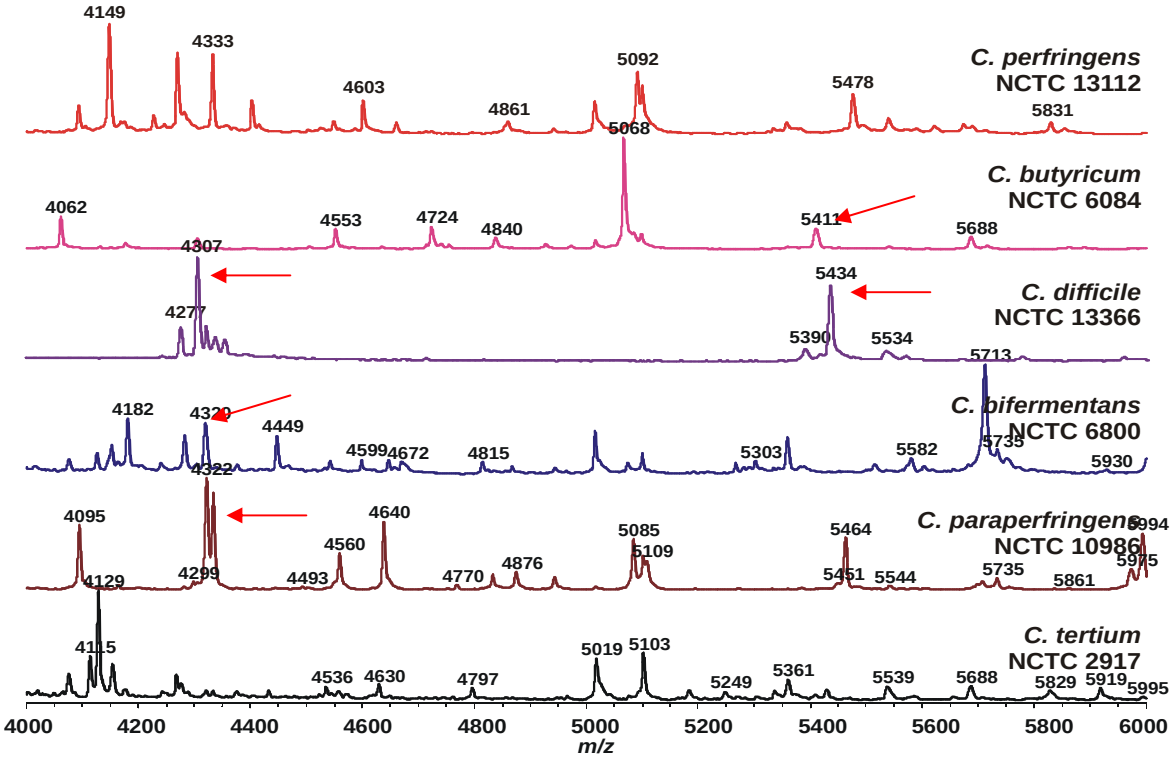


Figure 3.15: Mass spectral profiles of different species, closely related to *C. difficile*. Few common peaks could be seen among the species (red arrows).

3.4.2 Cluster analysis of the *C. difficile* isolates and closely related *Clostridium* species

All mass spectra obtained from the 24 h culture (duplicate spectra) were compared using the SARAMIS cluster analysis programme which is based on the percentage matching identical masses. All replicate spectra of individual isolates clustered closely. The incorrectly identified isolate MPRL 4857, clustered separately from the rest of the isolates. This was due to the contamination of the sample and poor spectra obtained from the MS analysis (Figure 3.16).

Cluster analysis was performed against different *Clostridium sp.* and clinical isolates of *C. difficile* formed separate clusters distant from the *C. difficile* isolates (Figure 3.17). Also, the duplicate mass spectra of different Clostridial species identified by the SARAMIS™ clustered together and each species formed a separate cluster (Figure 3.18).

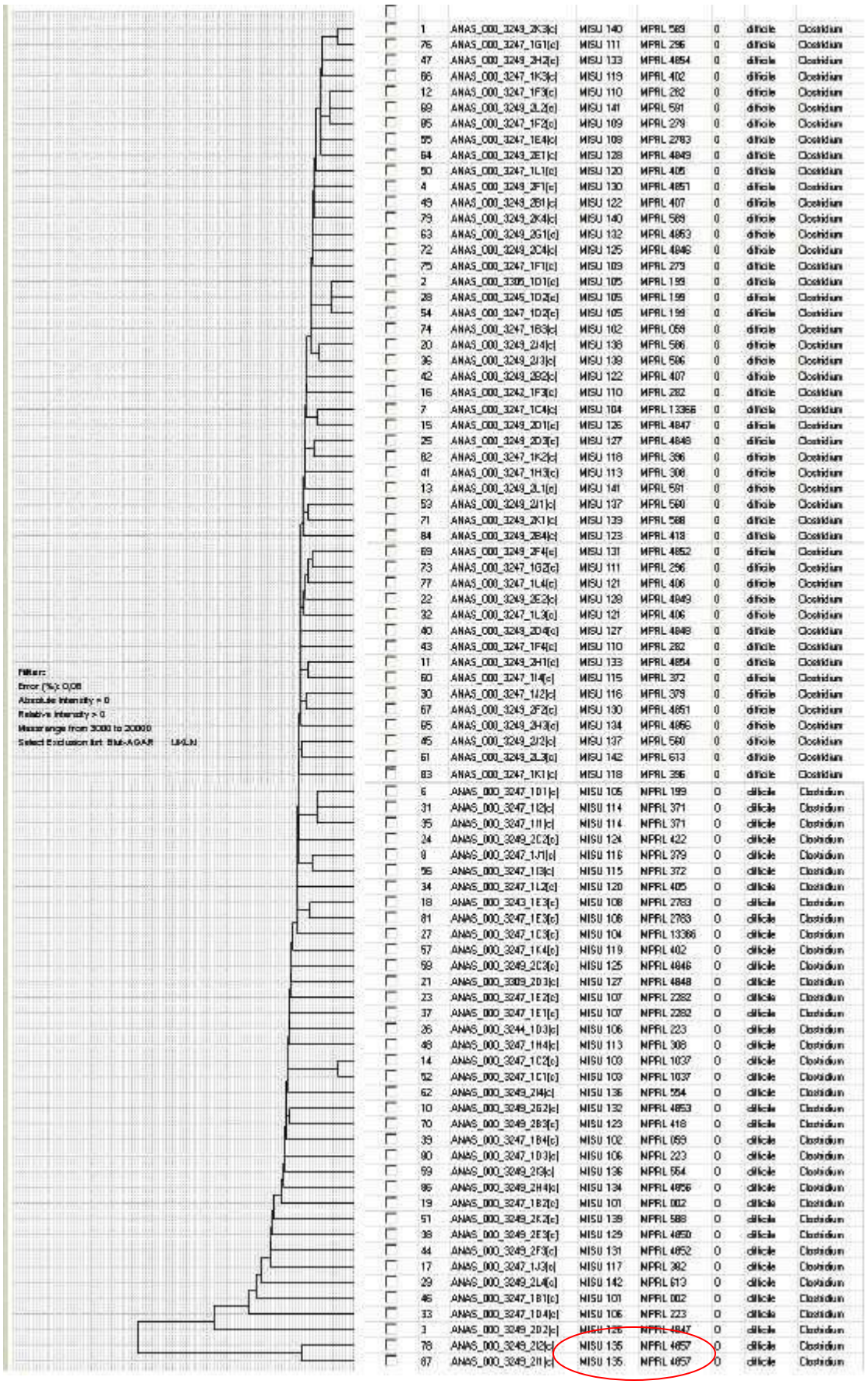


Figure 3.16: Cluster analysis of mass spectra in the mass range of 2,000 – 20,000 Da on cultures grown on CBA for 24 h. The clustering is based on mass signals detected in multiple samples. The last two spectra are from the isolate MPRL 4857 (red), identified as *S. warneri*.

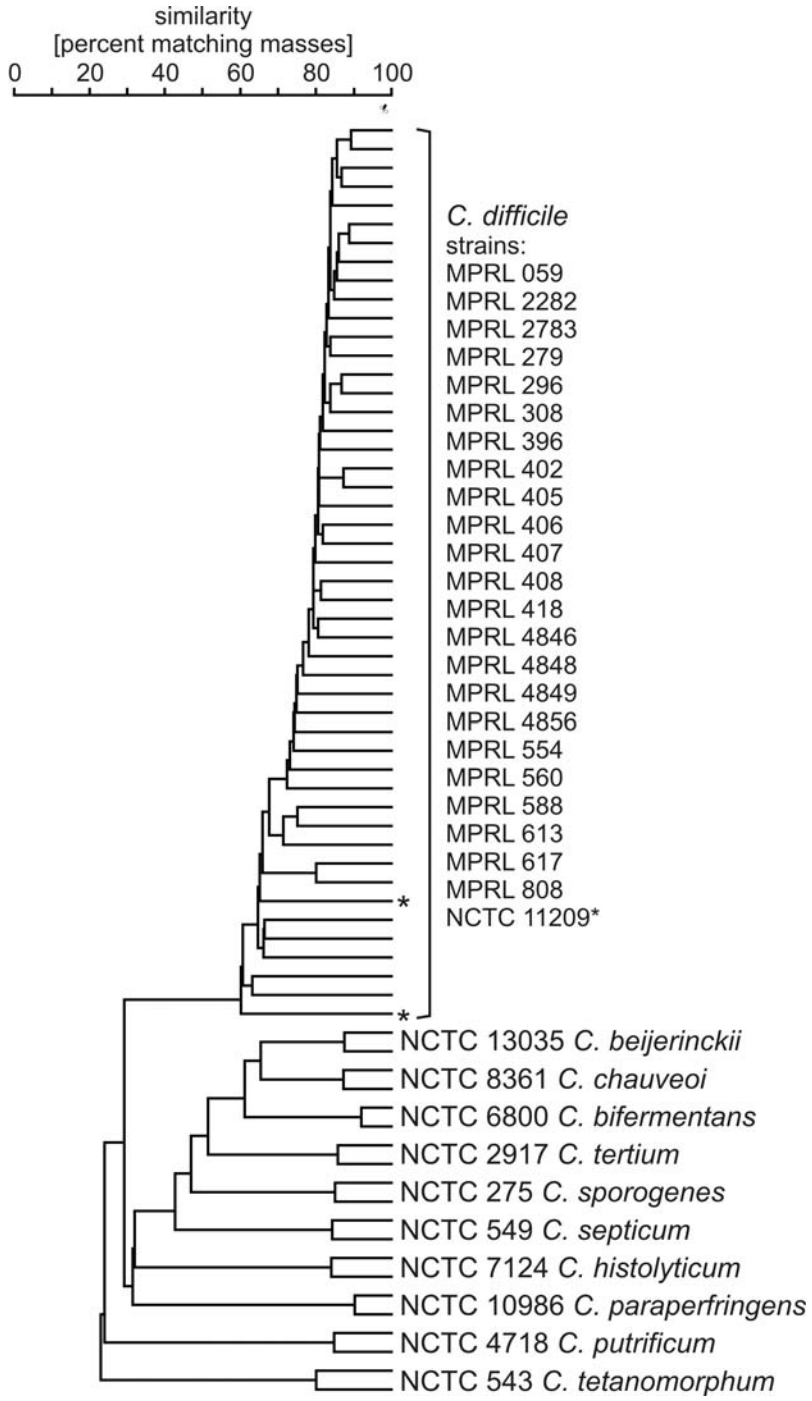


Figure 3.17: Cluster analysis of mass spectra in the mass range of 2000- 20000 Da. Differentiation of clinical isolates of *C. difficile* from other *Clostridium* species. The latter showed < 30 % similarity of mass ions. The Type strain of *C. difficile* is shown in *.

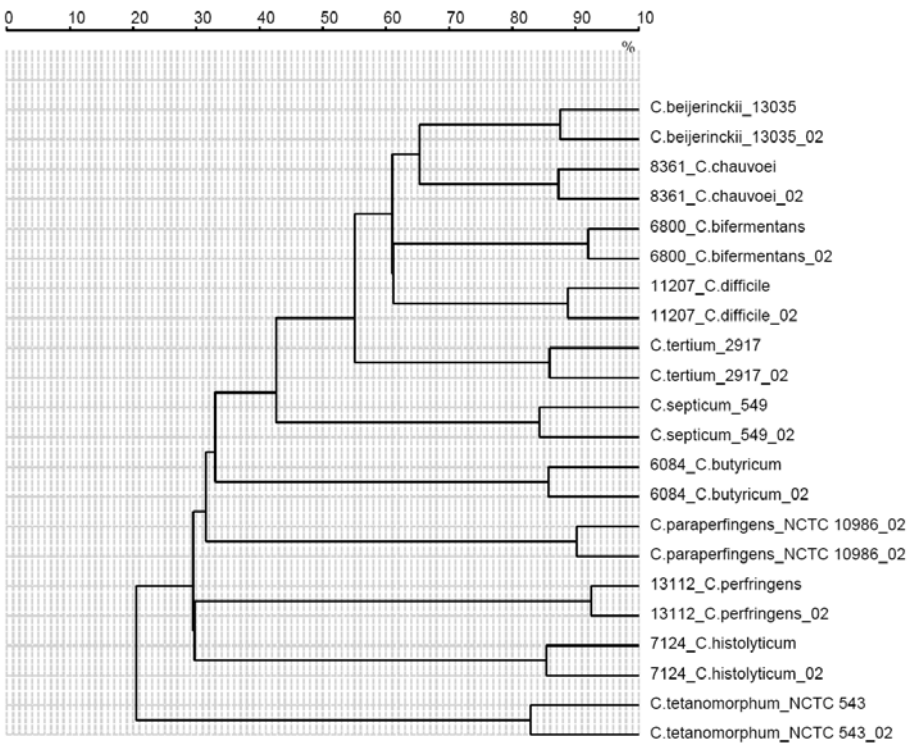


Figure 3.18: Cluster analysis of the MS profiles of different species of Clostridia analysed in duplicates.

3.5 SELDI-TOF-MS analysis

3.5.1 Optimisation of a protein extraction for *S. aureus*

In order to obtain sufficient protein for SELDI analysis, different protein extraction methods were attempted. Since *S. aureus* is a robust gram-positive organism, lysostaphin, an enzyme which helps to dissociate the cell wall components was used in the protocol. Protein extractions were carried out using a French press and Mickle beating combined with freeze thawing using a standard lysis solution and water resulted in high protein concentrations. When the samples were run on SELDI and NuPAGE® gels, both the methods showed similar results.

Spectral profiles of the samples prepared by using French press using distilled water showed a higher number of peaks than the samples prepared using lysis solution. The samples prepared using Mickle beating in the presence of lysis solution also showed a greater number of peaks with a higher intensity when compared to the other spectra (Figure 3.19). Comparison of the samples on NuPAGE® gels revealed that samples prepared using the French press had a higher number of protein bands. However, the sample prepared using Mickel beating in the presence of lysis solution also showed a similar result (Figure 3.20).

Comparison of the protein concentrations of the samples prepared using different concentrations of lysostaphin revealed the samples prepared using 30 µg of lysostaphin gave a higher protein yield. Hence for the rest of the study, protein extractions were carried out using the Mickle beating with lysostaphin in the presence of 30 µg of lysostaphin.

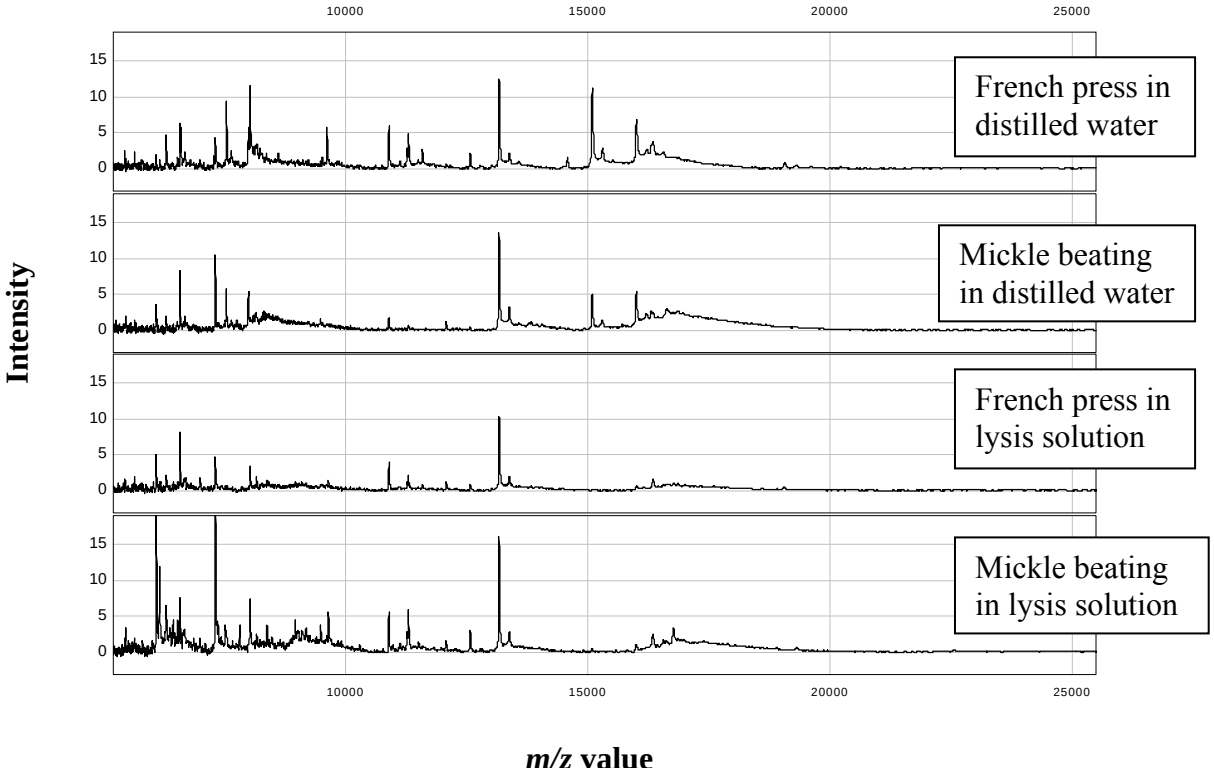


Figure 3.19: Comparison of the SELDI-TOF-MS profiles of MRSA 252 using the French Pressure cell and Mickle beating cell extracts with a protein concentration of 300 mg/ml. Extraction was performed both in distilled water and a standard lysis solution. The profiles obtained were very similar with the larger number of peaks present in the Mickle beating extract prepared in lysis solution.

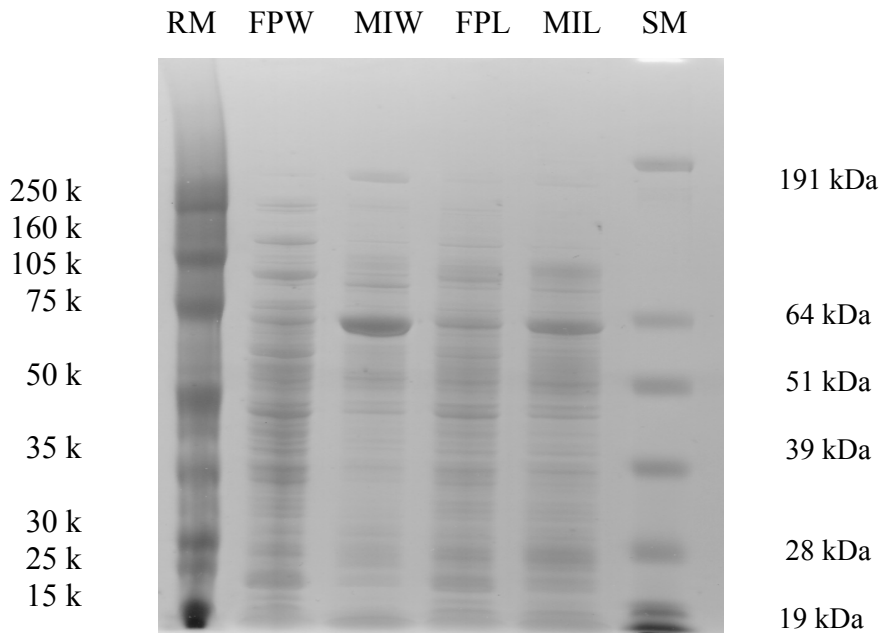


Figure 3.20: Comparison between the protein extractions using the French Press and Mickle beating. Each done by using a standard lysis solution and water. RM= Rainbow Marker, FPW= French Press using Water, MIW= Mickle beating using Water, FPL= French Press with Lysis solution, MIL= Mickle beating using Lysis solution, SM= See blue Marker. Total protein concentration was 1 µg/µl and 10 µl was loaded on each well.

3.5.2 Different ProteinChip® arrays tested for MRSA and MSSA

From the three surfaces investigated, there were 20 peaks detected in the Hydrophobic (H50) surface profiles, 34 in the Strong Anionic Exchange (SAX) profile and 42 in the weak cation exchange (CM10). The CM10 array also showed high intensity spectra with lower background noise and a several high-molecular weight peaks (20-30 kDa). The H50 profile contained mainly low intensity peaks while the SAX array displayed high level of background noise (Figure 3.21). Overall the CM10 chip produced more resolved peaks with less noise and was chosen for further analysis.

Using the CM10 ProteinChip® array, preliminary work carried out on MRSA and MSSA isolates showed that they both shared similar and reproducible profiles (Figure 3.22).

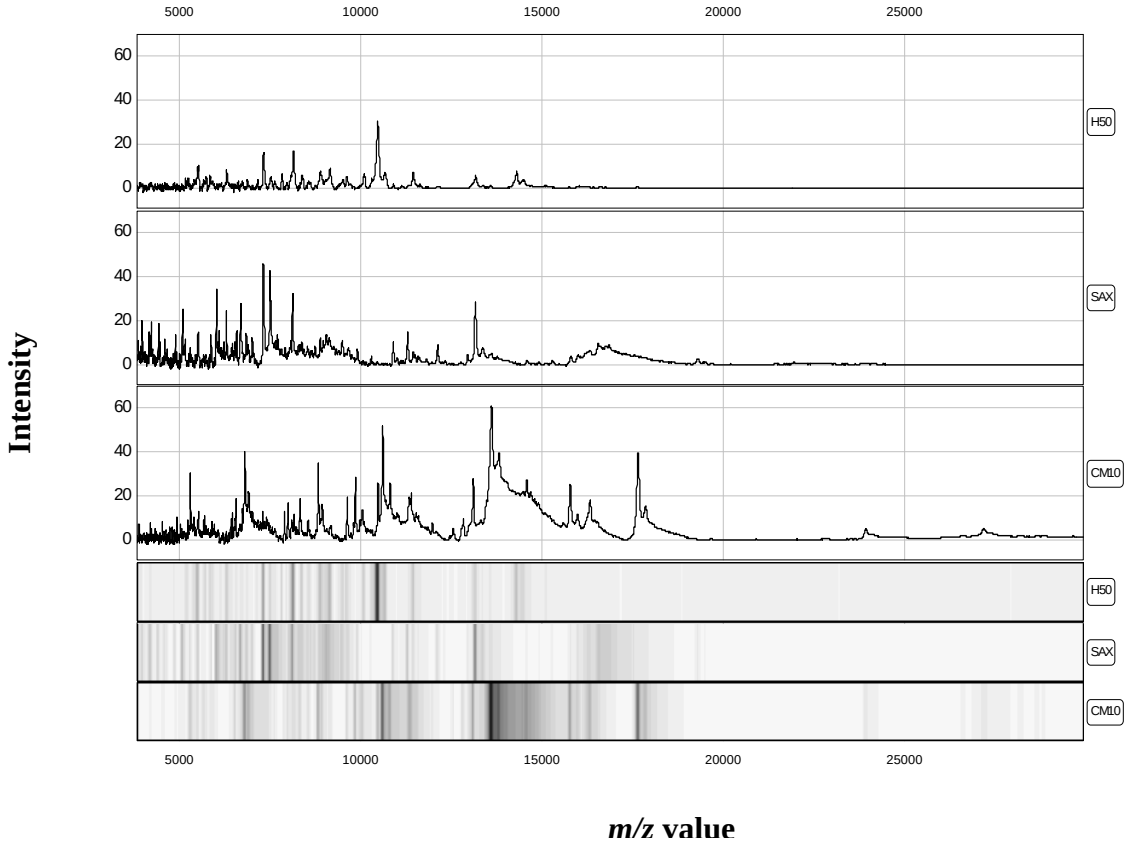


Figure 3.21: Comparison of H50, SAX and CM10 ProteinChips® for *Staphylococus aureus* (NCTC8325) on SELDI-TOF-MS with a protein concentration of 150 mg/ml on each spot. The CM10 array showed more resolved peaks with less noise compared to the other two surfaces.

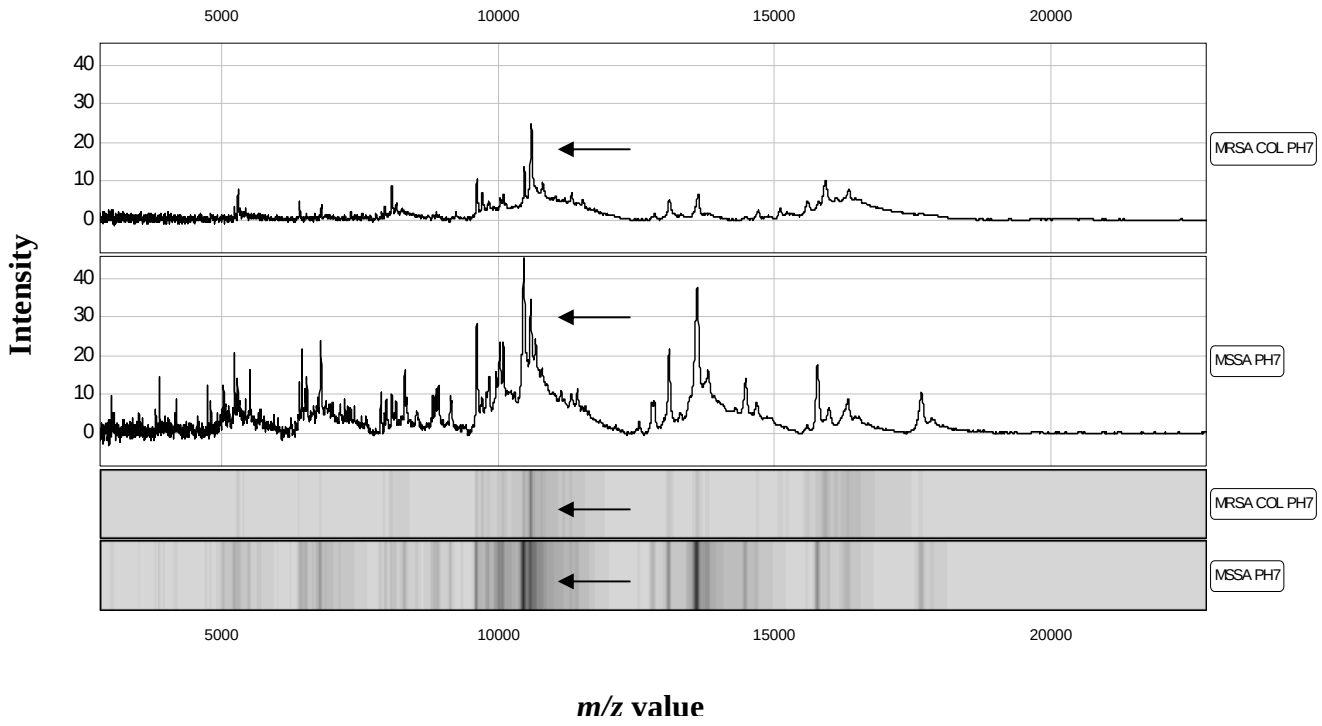


Figure 3.22: Comparison of MRSA (COL) and MSSA (NCTC 8325) on SELDI-TOF-MS using CM10 with a protein concentration of 150 mg/ml on each spot, intensity of 200. The gel view representation of the profile clearly shows the common peaks present in both MRSA and MSSA.

3.5.3 Effect of pH on CM10 ProteinChip® arrays for MRSA and MSSA

Comparing the spectra from pH = 4, 6, 7 and 8, all four spectral profiles gave similar patterns. But the spectral profile of the pH = 7, gave a slightly more stable profile compared to the other pHs with high intensity peaks (Figure 3.23). For the rest of the study pH = 7 was used when preparing the buffer.

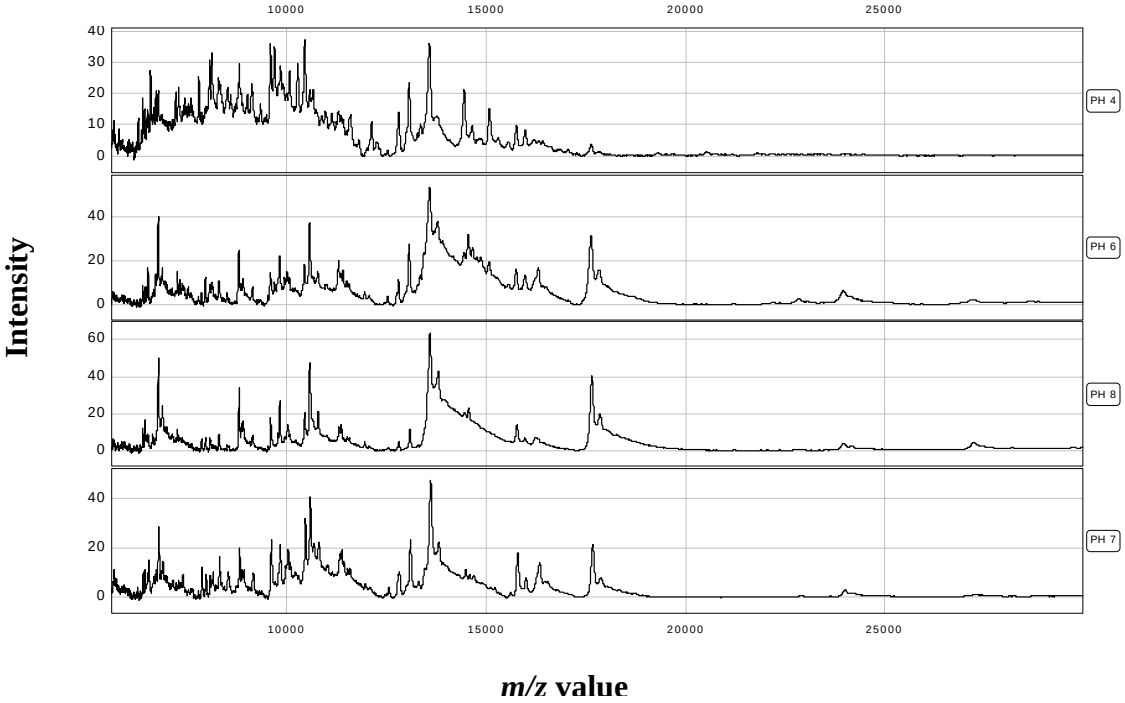


Figure 3.23: Effect of pH on CM10 array tested using *S. aureus* NCTC 8325. Compared to pH = 4, 6 and 8, pH = 7 spectral profile showed a higher number of peaks and had a more stable profile.

3.6 Artificial Neural Networks (ANN) analysis of MRSA and MSSA

SELDI-TOF-MS profiles were obtained from a total of 99 *S. aureus* isolates where 49 (n = 49) of which were MRSA and the rest were (n = 50) MSSA. However, the complex nature of the SELDI data and the large number of spectra accumulated made the analysis difficult to perform using comparative gel view images and heat maps. It was evident that a single ion could not be used to distinguish between MRSA and MSSA and a combination of several of biomarkers was required.

Ion mass intensity profiles generated from SELDI analysis were analysed in a stepwise approach in order to rank the ions based on the ability to predict MRSA and MSSA. From the first step, one key ion (3,081 Da) which was able to predict MRSA compared to the other ions, was chosen for the next stepwise analysis. This kind of model parameterisation steps ensued with the aim of identifying which ions are the most important within the dataset, allowing the model complexity to be reduced and the predictive capabilities to be increased. The chosen ion 3,081 Da was then run through 50 sub-models including the rest of the ions to determine the most significant sub-set of ions, capable of predicting MRSA and MSSA.

After training the model with all the ions including the best chosen ion in order to determine the best subset of ions, seven ions were predicted as key ions within the m: z range of 3,000 Da – 19,000 Da and chosen for further analysis. This reduction is necessary in order to find any potential biomarkers. However, when calculating the mean error for all the seven ions, for each sub-set of ions the mean error drops, and at one point (18,896 Da) the error change becomes insignificant indicating that, beyond this point there were no significant predictive ions present. Hence, the seven most predictive ions (3,081 Da, 5,709 Da, 5,893 Da, 7,694 Da, 9,580 Da, 15,308 Da and 18,896 Da) were chosen to recreate the model (Figure 3.24 and 3.25).

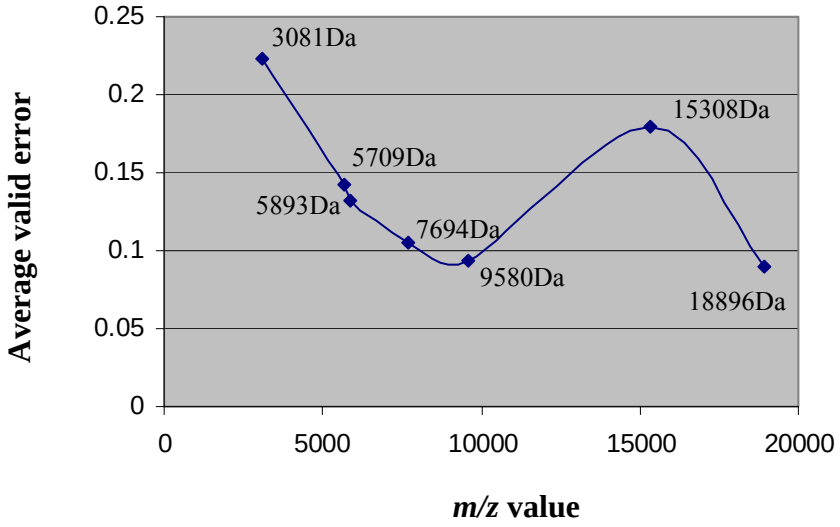
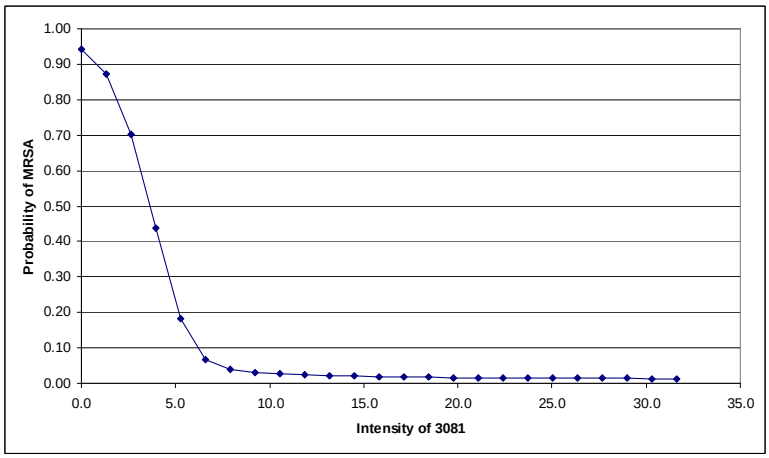
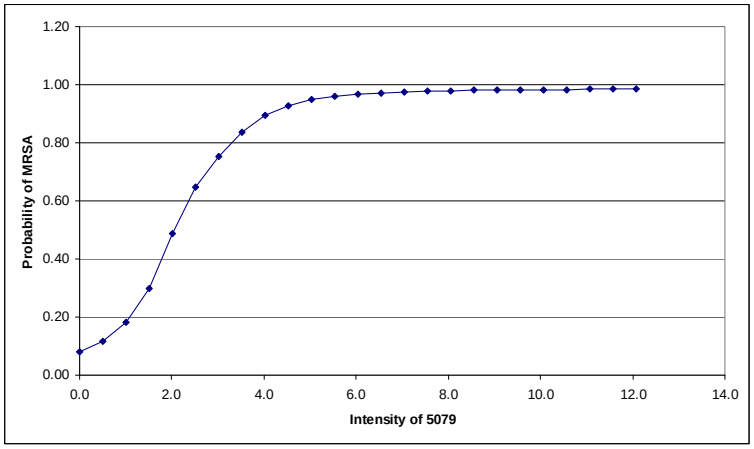


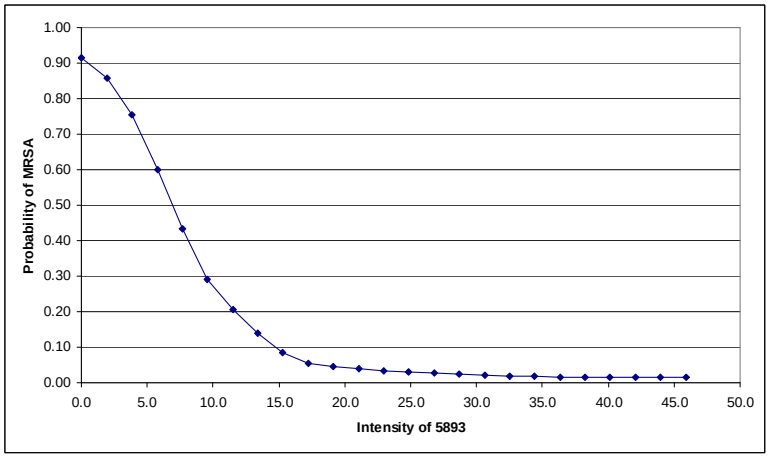
Figure 3.24: Stepwise approach of choosing the most significant sub-set of ions, which can be used to predict MRSA and MSSA. After the seventh ion 18,896 Da, mean error variation becomes insignificant and chosen as the best sub-set of ions for predicting MRSA and MSSA.



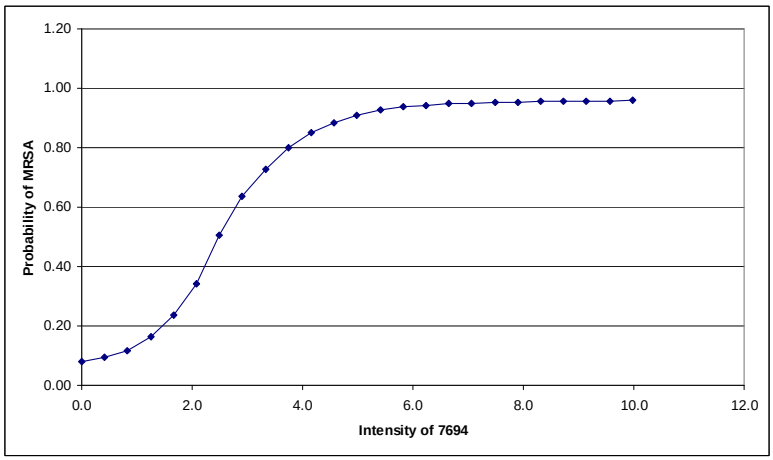
(a) 3081 Da



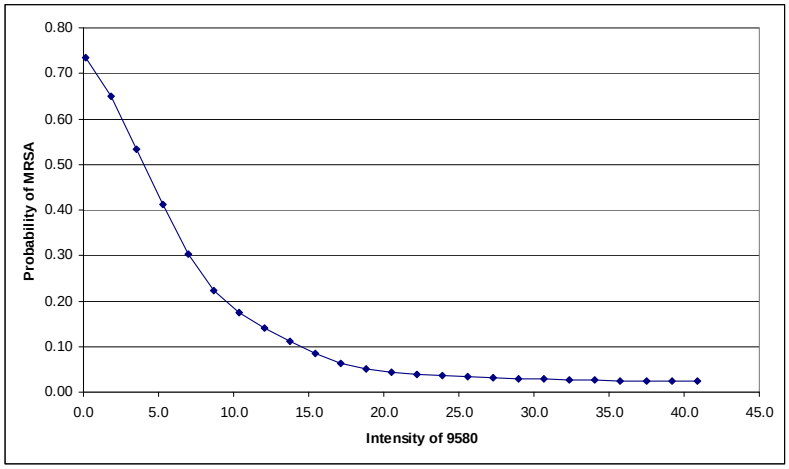
(b) 5709 Da



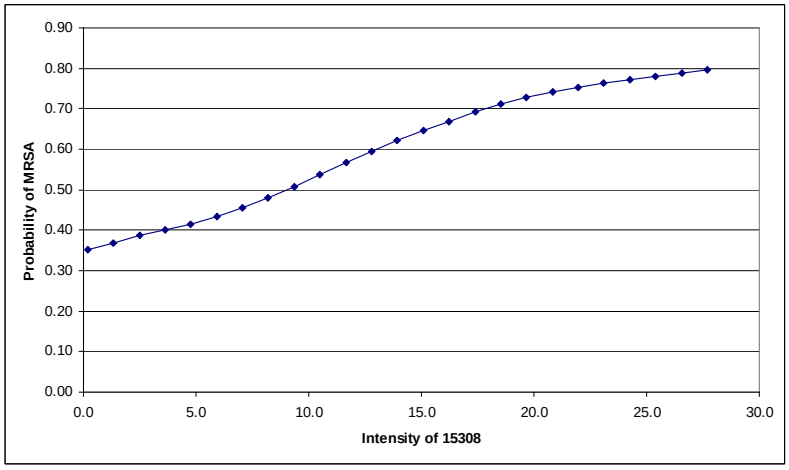
(c) 5893 Da



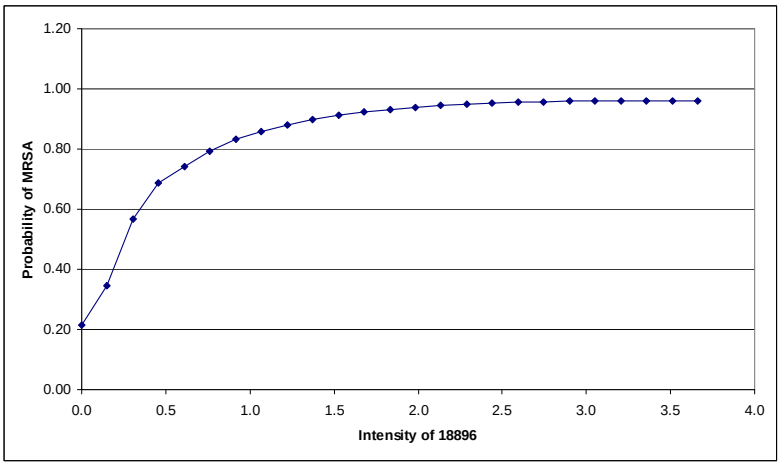
(d) 7694 Da



(e) 9580 Da



(f) 15308 Da

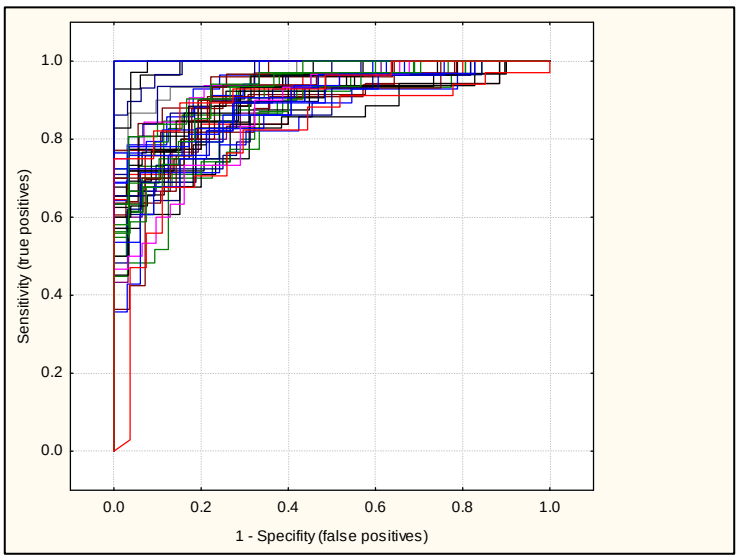


(g) 18896 Da

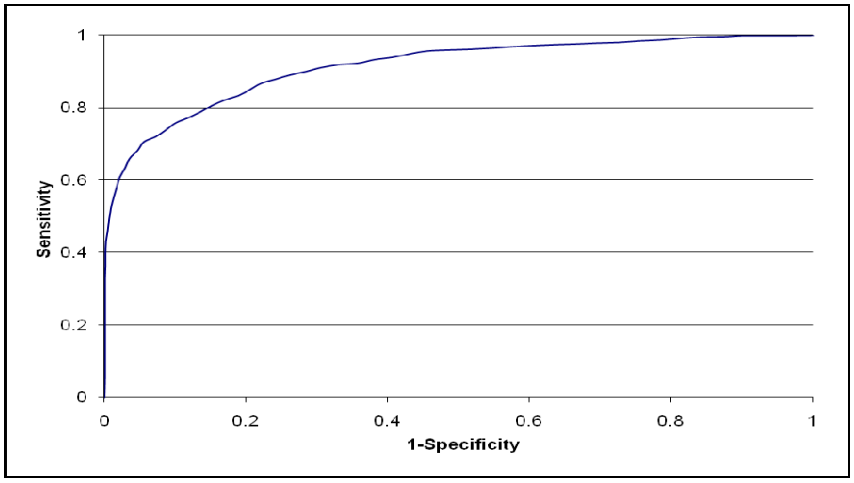
Figure 3.25: Seven key mass ions chosen for the prediction of MRSA and MSSA showing the capability of predicting with the increasing of the intensity. In ions 3081 Da (a), 5893 Da (c) and 9580 Da (e), predictive performance decreases with the increasing intensity while 5709 Da (b), 7694 Da (d), 15308 Da (f) and 18896 Da (g) ions performance increase with the intensity.

From the population distribution curve of MRSA and MSSA, most of the MRSA isolates were correctly classified as MRSA and only two MSSA isolates were incorrectly predicted by the key ions (Figure 3.27). When the isolates reached closer to the predictive value of 1, most of the MRSA isolates were predicted 100 % as MRSA indicating that the seven ions chosen for the prediction are characteristic to MRSA. The same pattern could be seen for the MSSA isolates and the misclassified isolates in the middle, which probably represent strains that are in the process of gaining antibiotic resistance.

The Receiver operating curve (ROC) from this seven ion input model showed an overall area under the curve (AUC) value of 0.9147, indicating that the seven ions chosen for the prediction of MRSA and MSSA in combination have high sensitivity and specificity (91%) (Figure 3.26). However, a new blind dataset could be used to validate the model further.



(a)



(b)

Figure 3.26: Receiver operating curves (ROC) and Area under the curve (AUC) values for the first 50 ANN sub-models (a). Mean AUC of all 50 sub-models (b) indicates a high sensitivity and specificity (91%).

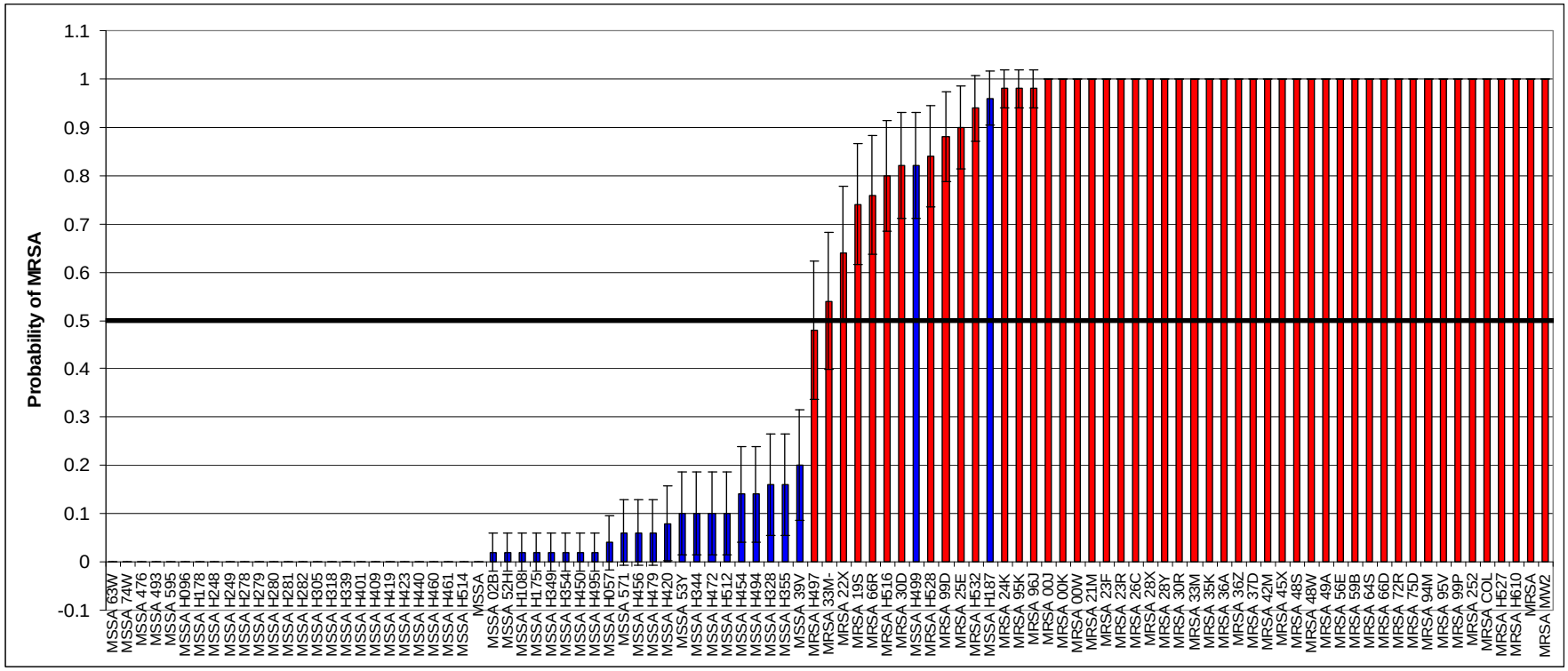


Figure 3.27: Population distribution of MRSA and MSSA. The red colour bars indicate MRSA while the blue colour bars indicate MSSA strains.

3.7 SDS-PAGE analysis of MRSA and MSSA

From the two extraction methods investigated, results from the boiling method gave a very poor protein concentration and was not suitable for further analysis. The protein extracts prepared using lysostaphin gave a higher protein concentration suitable for 1D gel electrophoresis and was chosen for further studies.

A pilot study with a total of 8 MRSA and MSSA isolates including type/reference isolates were run using precast NuPAGE® gels to visualise any differences between strains (Figure 3.28). Most of the proteins were conserved within the molecular weight range of 39 - 64 kDa and 19 - 28 kDa. Although the proteins are conserved within that region, one of the MSSA isolates (5) showed a different profile to the MRSA strain. These gel-to-gel comparison of the isolates showed that there are differences between MRSA and MSSA and further MS/MS studies on these isolates could help to identify unique proteins.

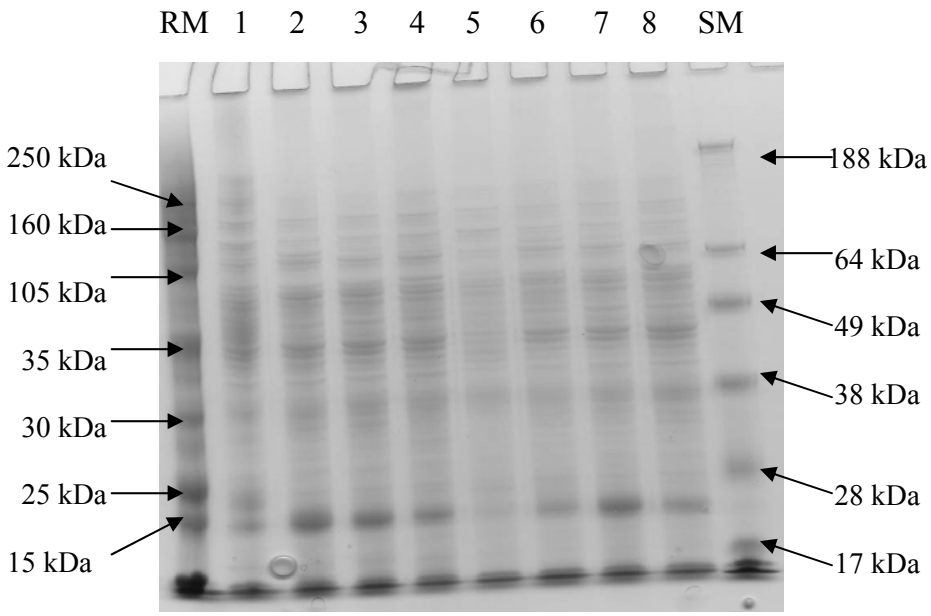


Figure 3.28: Comparison of MRSA and MSSA using NuPAGE® Bis-Tris precast polyacrylamide Gels:

RM=Rainbow marker, 1=MRSA 88, 2=MRSA 25E, 3=MRSA MW2, 4=MRSA COL, 5=MSSA 53Y, 6=MSSA O2B, 7=MSSA NCTC 8325, 8=MSSA 476, SM=See blue Marker. Total protein concentration was 1µg/µl and 10 µl was loaded on each well. Protein bands could be observed in the molecular weight range of 39 - 64 kDa and 19 - 28 kDa.

3.8 SELDI-TOF-MS and SDS-PAGE analysis of *C. difficile*

3.8.1 Selection of a protein extraction method for *C. difficile*

With the protein extraction method described above, (2.2.1), *C. difficile* yielded a protein concentration of ~ 5.0-10.0 µg/µl. The protein concentrations did not improve with the addition of lysozyme and was omitted in future work. Since no further optimisation was required for the extraction method, using this protocol SELDI-TOF-MS analysis was carried out on three different ProteinChip® arrays.

A total of 114 (n = 114) isolates were run on SELDI and 28 (n= 28) isolates were analysed using SDS-PAGE gels.

3.8.2 Different ProteinChip® arrays used in the study for SELDI analysis

The weak cationic exchange (CM10), Hydrophobic (H50) and Strong Anionic Exchange (SAX) were used to analyse the *C. difficile* extracts. Both weak cationic exchange and hydrophobic chips produced six peaks while the Strong Anionic Exchange ProteinChip® array detected about 21 peaks. Since the profiles obtained using the Strong Anionic Exchange ProteinChip® array (SAX/Q10) showed a larger number of peaks with higher intensity, the Q10 array was chosen for further studies (Figure 3.29).

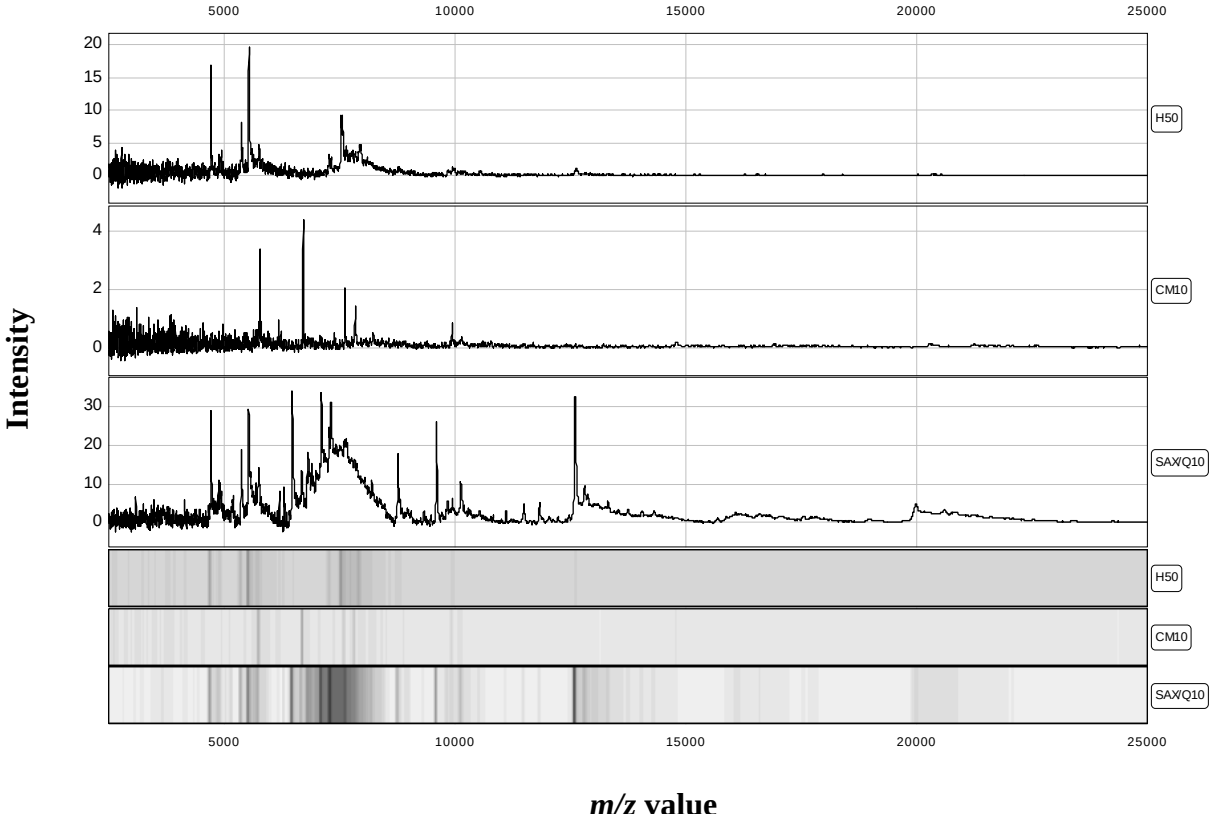


Figure 3.29: Comparison of H50, CM10 and SAX/Q10 ProteinChips® mass spectral profiles of *Clostridium difficile* (MPRL 47). The protein concentration of 300 mg/ml was applied to each spot. SAX/Q10 array showed more resolved peaks with less noise compared to the other two surfaces.

Characteristic peaks/mass ions were obtained for all cell extracts but different peaks were observed for some isolates indicating that changes have occurred among isolates over the years (Figure 3.30 and 3.31).

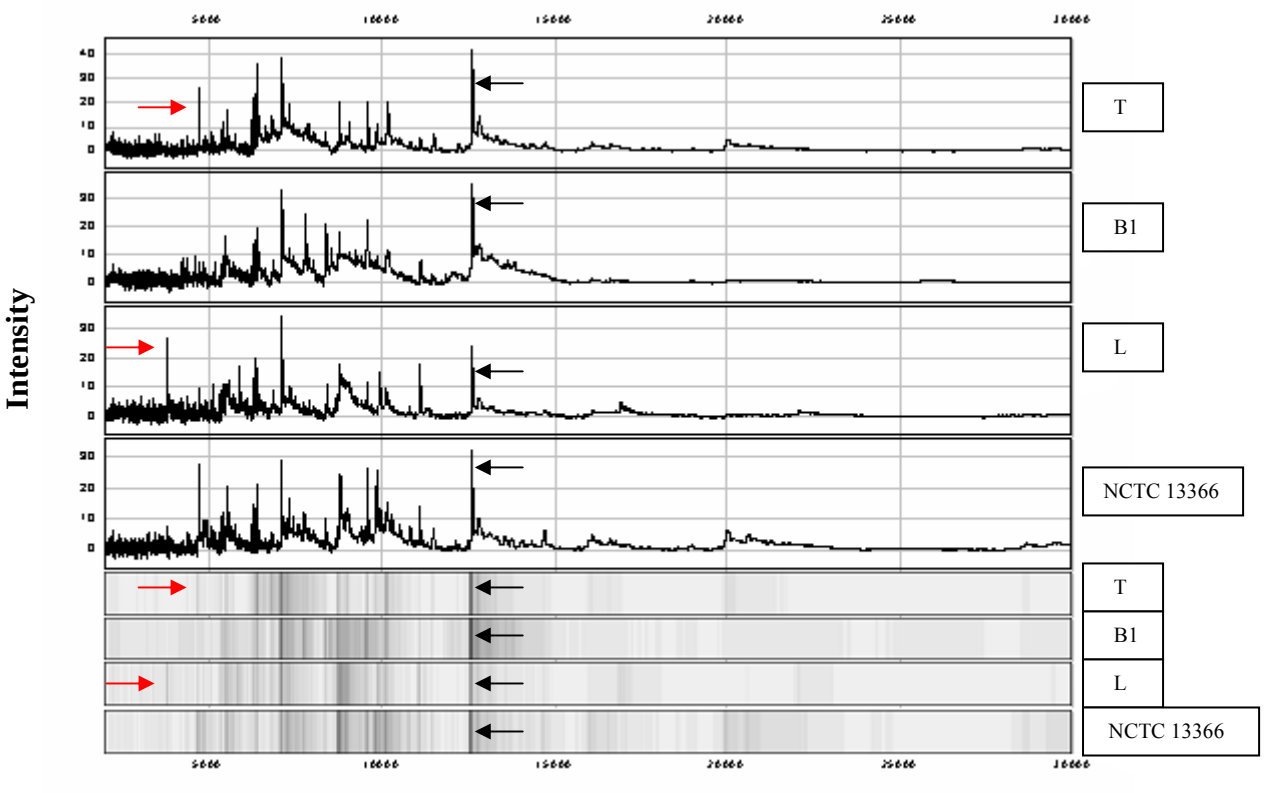


Figure 3.30: Comparison of three *C. difficile* isolates T, B1 and L against the reference *C. difficile* strain NCTC 13366. Common mass ions (Blue arrows) were present among the four isolates but there is clear evidence of intraspecies diversity. For example, additional strain specific mass ions were present in T and L strains (Red arrows).

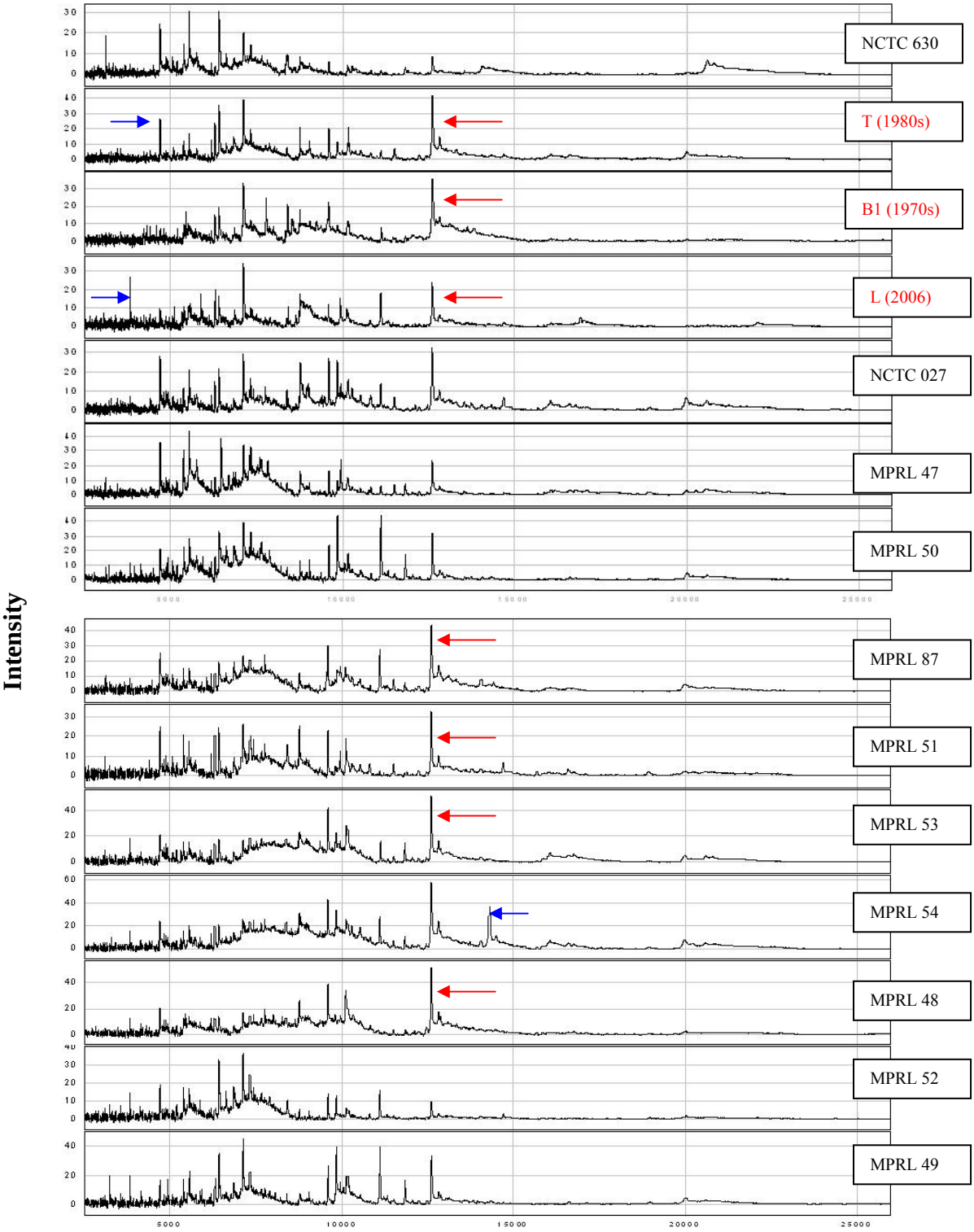


Figure 3.31: A cross section of SELDI-TOF-MS profiles of 14 isolates of *C. difficile* using SAX/Q10 ProteinChip® arrays, including the test isolates T, B1 and L. Common mass ions (red arrows) could be seen among all the strains while strain specific peaks were also detected (blue arrows).

3.8.3 One dimensional gel analysis of *C. difficile*

To visualise protein patterns between isolates, the 28 isolates were run on 1D gel (NuPAGE®) having previously adjusted all extracts to the same protein concentration. Although the profiles were similar, most of the proteins were conserved in the ~ 38 - 49 kDa region while few showed bands in the molecular weight range of 6 kDa and 17 kDa (Figure 3.32). The reference isolates L, T and B1 were also similar to the rest of the isolates.

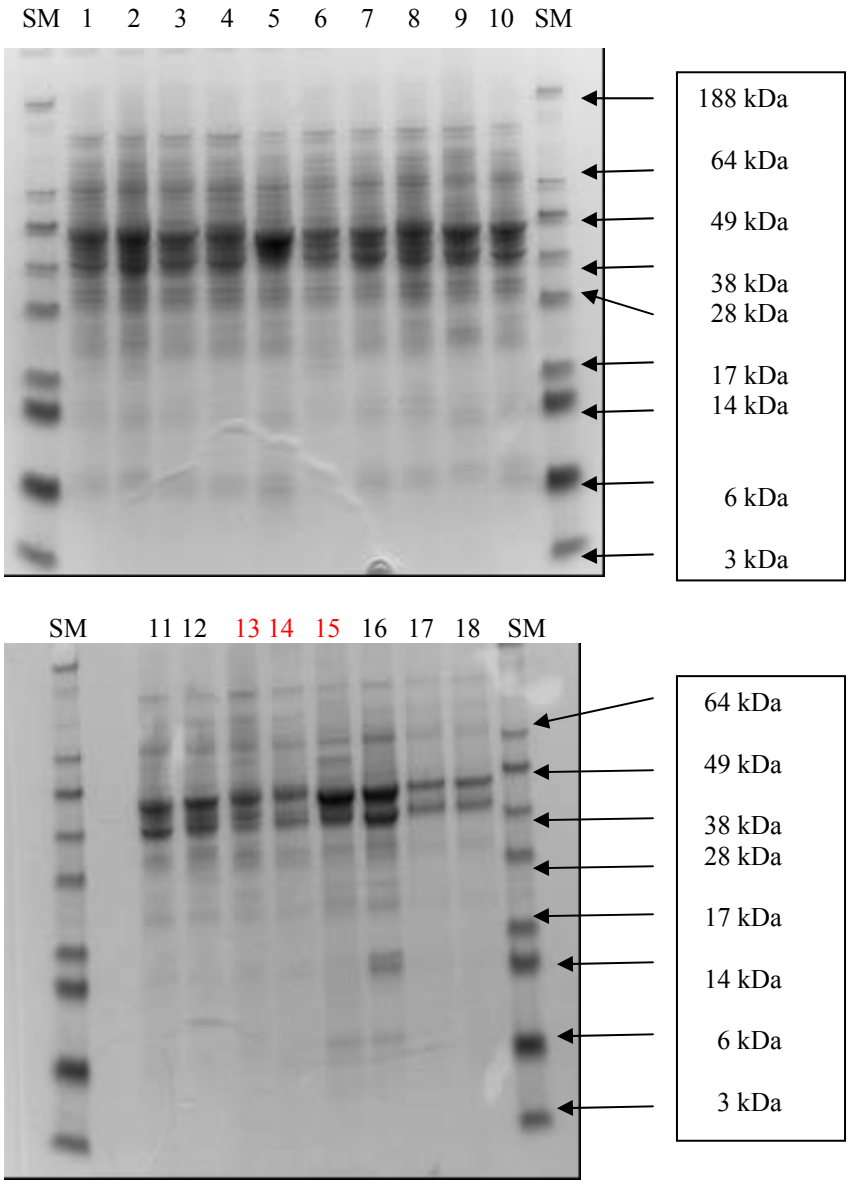


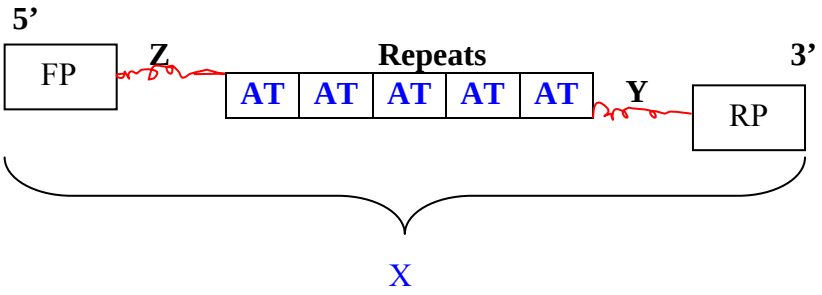
Figure 3.32: Comparison of the SDS-PAGE profiles of 18 *C. difficile* isolates. Top gel: 1-10: NCTC 13287, NCTC 630, NCTC 027, NCTC 106, MPRL 47, MPRL 48, MPRL 49, MPRL 50, MPRL 51 and MPRL 52 respectively. Second gel: 11-18: MPRL 616, MPRL 1037, T, L, B1, MPRL 613, MPRL 2783 and MPRL 808 respectively. SM= See blue Marker. Overall the band patterns were homogenous with the exception, several unique bands were present in the profiles of MPRL 613 (16) and B1 (15) observed in the low molecular weight area ~ 6 kDa - 17 kDa.

3.9 VNTR analysis of *C. difficile*

3.9.1 VNTR copy number calculation

A total number of 47 loci were tested for this study. Ten of these were published earlier (van den Berg *et al.,* 2007 and Marsh *et al.,* 2006). Tandem Repeats Finder software was used to identify the repeat elements using the sequenced genome of *C. difficile* 630 strain as reference. Primers for the 37 loci were designed using Primer3 programme by selecting suitable primers from the conserved regions flanking the repeats at the 5' and 3' end.

To calculate the copy number of a particular locus, the total number of nucleotides in the flanking regions were deducted from the size of the amplified product and divided by the repeat size (the number of bases present in that particular repeat). The copy number was calculated manually as shown below (Figure 3.33). For each isolate, this number was imported into BioNumerics software for comparison and was used to detect the variability within the same ribotype as well as the variability among the different ribotypes.



Forward primer: FP

Reverse primer: RP

Product size: X

Flanking regions: Z and Y

$$\text{Copy number} = \frac{(X) - (Z+Y)}{\text{Repeat size}}$$

Figure 3.33: Formula used to calculate the copy number of a particular locus. The amplified product size is indicated as X and the flanking regions are labelled as Z and Y while the forward and reverse primers were labelled as FP and RP respectively.

3.9.2 VNTR analysis

A total of 92 *C. difficile* isolates were analysed for this study and 47 VNTR loci were selected as described in methods (section 2.8.2 and 2.8.3). DNA was extracted for each isolate using the MagNa Pure LC Robot. Multiplex PCRs were carried out for each isolate. Null results (i.e. absence of amplified product) in isolates were confirmed by carrying out reactions with only one primer pair in the PCR.

Of the 47 loci, seven loci (CD16, 17, 26, 29, 31, 44 and 47) failed to give an amplification product from any isolate in the set. Hence, these loci were excluded from analysis. Also primer CD9, which was previously published, was unable to amplify repeats from any isolate. The remaining 39 primer pairs, including the published primers, successfully amplified the repeats for most of the *C. difficile* isolates (Tables 3.4, 3.5 and 3.6). For few isolates, although the primers were able to amplify and generate an amplicon, the size was identical to the sum of the two flanking regions. This indicated the absence of repeats in that isolate i.e. copy number was '0'. In some isolates, no product was amplified. These results were referred to as 'null set' (Tables 3.4, 3.5 and 3.6). The copy number for each isolate was calculated manually as described above.

3.9.3 Cluster analysis using BioNumerics

In order to understand and differentiate the strains of different ribotypes using the VNTR data, the bioinformatics tool, BioNumerics version 5.1 (Applied Maths, Belgium) was used. The copy numbers obtained for each isolate for different loci were imported into the BioNumerics database. Isolates that gave 'null' results i.e. no amplification product were entered as '99' into the database as BioNumerics does not accept alphabets to be entered into its database. Also the copy numbers were rounded to the nearest integer when entering into the database e.g. 3.3 = 3 and 3.6 = 4. Cluster analysis was performed using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) algorithm for all the isolates. Parameters such as similarity coefficient, Dice, was used for the cluster analysis which statistically compare the similarity or diversity of two sample sets.

The stability tests performed on VNTR loci concluded that isolates with summed tandem repeat differences of ≤ 2 are genetically related (Marsh *et al.,* 2006) but differences of one repeat unit between isolates should not be interpreted as separate subtypes (van den Berg *et al.,* 2007). Another study using epidemiological data analysis of *C. difficile* isolates concluded that a difference of up to two repeat units (i.e. > 71% similarity) indicates closely related strains (Fawley *et al.,* 2008). This indicates that even though there are one or two copy number differences within a locus among the isolates, they could be referred to as closely related isolates.

VNTR Locus	Location	Size (bp)	Copy no. range 027
CD 1	755721-755950	6	12-36
CD 2	3688632-3688751	7	11-46
CD 3	3239736-3239835	6	14-50
CD 4	167124-167172	7	9-12
CD 5	1954913-1954939	3	5
CD 6	664660-664705	8	6-17
CD 7	4116072-4116109	9	2
CD 8	692929-693015	8	3-14
CD10	677132-677386	17	0-18
CD11†	2168250-2168268	9	1
CD12†	804519-804537	9	2
CD13†	301056-301074	3	5
CD14	788448-788465	9	Null
CD15†	800026-800043	9	2
CD18†	3014724-3014741	9	2
CD19†	2167938-2167955	9	2
CD20	2169070-2169086	7	2*
CD21†	2169418-2169435	9	2
CD22†β	804623-804645	3	1
CD23†	804652-804699	18	0
CD24	478055-478070	7	2*
CD25† β	543601-543619	9	2
CD 27	3246180-3246195	6	2*
CD 28† β	3246752-3246770	9	2
CD 30† β	794475-794493	7	2
CD 32† β	252622-252644	6	3
CD 33†	800719-800741	2	10
CD 34† β	799700-799717	6	6
CD 35† β	797131-797146	7	1
CD 36† β	881618-881637	10	2
CD 37† β	881702-881718	8	1
CD 38† β	883903-883918	6	2
CD 39† β	884179-884194	6	2
CD 40† β	884601-884612	3	4
CD 41 β	886202-886220	9	2*
CD 42 β	923152-923190	20	19*
CD 43	623650-623691	21	20-21*
CD 45	3319803-3320164	45	0-4*
CD 46†	3753183-3753574	45	6

Table 3.4: Summary of the VNTR copy numbers for ribotype 027 isolates for 39 loci exhibited by majority of isolates. Locus CD23 was observed with a copy number of '0'. For loci CD12, 14 and 46 (shown in red) a few isolates also produced copy numbers of Null, 2 and Null respectively.

* indicates majority of the isolates produce the same copy number and only a few isolates exhibit one or two different copy numbers.
† indicates stable loci (22 loci) with identical copy numbers for most of the isolates.
β indicates the loci (14 loci) excluded from the first cluster analysis.

3.9.4 Cluster analysis of ribotype 027 isolates

Fourteen of the 39 loci exhibited identical copy numbers for all 45 ribotype 027 isolates. A further 22 loci showed identical copy numbers for most of the ribotype 027 isolates (Table 3.4). Another seven loci exhibited least diversity (*viz* CD20, 24, 27, 41, 42, 43 and CD45) where only few isolates showed variable copy numbers compared to the rest of the ribotype 027 isolates. For the locus CD23, the copy number of the repeats was calculated as zero while the other ribotype strains showed a copy number of one, indicating CD23 as a marker for ribotype 027 to differentiate from strains of other ribotypes. The copy number for the locus CD14 was detected with a 'null' result for most of the isolates while only three isolates gave a copy number of two (Table 3.4). Hence this locus would also be a possible marker to differentiate ribotype 027 isolates. For loci, CD12 and CD46 the majority of the isolates gave copy numbers of two and six respectively, while few isolates were detected with null result. In general, for most of the loci, the presence of null results could be observed throughout ribotype 027 isolates. The null results were verified by repeating the VNTR PCR using single primer pair.

The most diverse VNTR loci with differential copy numbers for most of the isolates were CD1, 2, 3, 4, 6, 8, and, 10 (Table 3.4). These loci were identified earlier by Marsh and Van den Berg (van den Berg *et al.,* 2007 and Marsh *et al.,* 2006). According to Van den Berg's study, 3 of the 10 loci produced identical copy numbers for all ribotype 027 isolates; these loci were CD4 (copy number-10), CD5 (copy number- 4) and CD7 (copy number- 2). However in this study, only one locus, CD7 exhibited the same copy number (2) for all the 027 isolates. Also the locus CD5 exhibited an identical copy number (5) for all the isolates but different from the published copy number (4). The third locus, CD4, showed variable copy numbers in this study (9-12).

When performing cluster analysis, 14 loci which exhibited identical copy numbers for all the ribotype 027 isolates or loci which showed variable copy numbers in only one or two isolates were excluded from analysis (Figure 3.34). Also an arbitrary value of 90 % was considered to detect the similarity between the isolates. Eight clusters were detected among the isolates showing ≥ 90 % similarity within the isolates (A-H) (Figure 3.34). Also copy number differences were observed for isolates within each group to compare the stability test results by Marsh and van den Berg.

Comparing the copy number variability of isolates within each group, group A, consisted of two isolates which varied in the copy number of locus CD23; isolate G67 had '0' and G77 null copy number. Group B consisted of nine isolates, and seven of these isolates showed 100 % similarity to each other. The other 2 isolates i.e. G71 and G32 showed different copy numbers in loci CD27 (1) and CD23 (null) while the rest of the seven isolates showed copy numbers of two (2) and zero (0) for CD27 and CD23 respectively. In group C, out of the two isolates, G19 and G20, copy numbers varied in CD30 (null) and CD43 (20). In group D, isolate G31 showed different copy numbers in loci CD13 (null), CD27 (1) and CD33 (10) while isolate G35 showed copy numbers of five, two and 'null' for respective loci. Group E consisted of four isolates, and two of these exhibited 100 % similarity. The remaining two isolates, G09 and G41 showed different copy numbers in loci CD23 (1), CD14 (2) and CD46 (6) respectively. Four isolates were clustered together in group F and all these isolates showed 100 % similarity to each other. Group G consisted of two isolates including the sequenced strain 630, NCTC 13366, and G45. These two isolates showed different copy numbers in loci CD24 (1) and CD15 (null) respectively. Also both the isolates showed two different copy numbers in loci CD46 (null and 6 respectively). In the last group H, three isolates were detected with variable copy numbers in three loci. These were CD14 (null), CD23 (null) and CD24 (1).

Considering this variation in the copy numbers within the loci of each group, most of the variability observed was either with one or two copy number differences. Also in some cases, within a group, isolates either showed copy numbers of 'null' or '0' and 'null' or '6' for a particular loci, e.g. group G and B. This was because either there was no product amplified or the repeats were absent. Hence, all these isolates within each group could be considered as genetically related although there were one or two copy number variances.

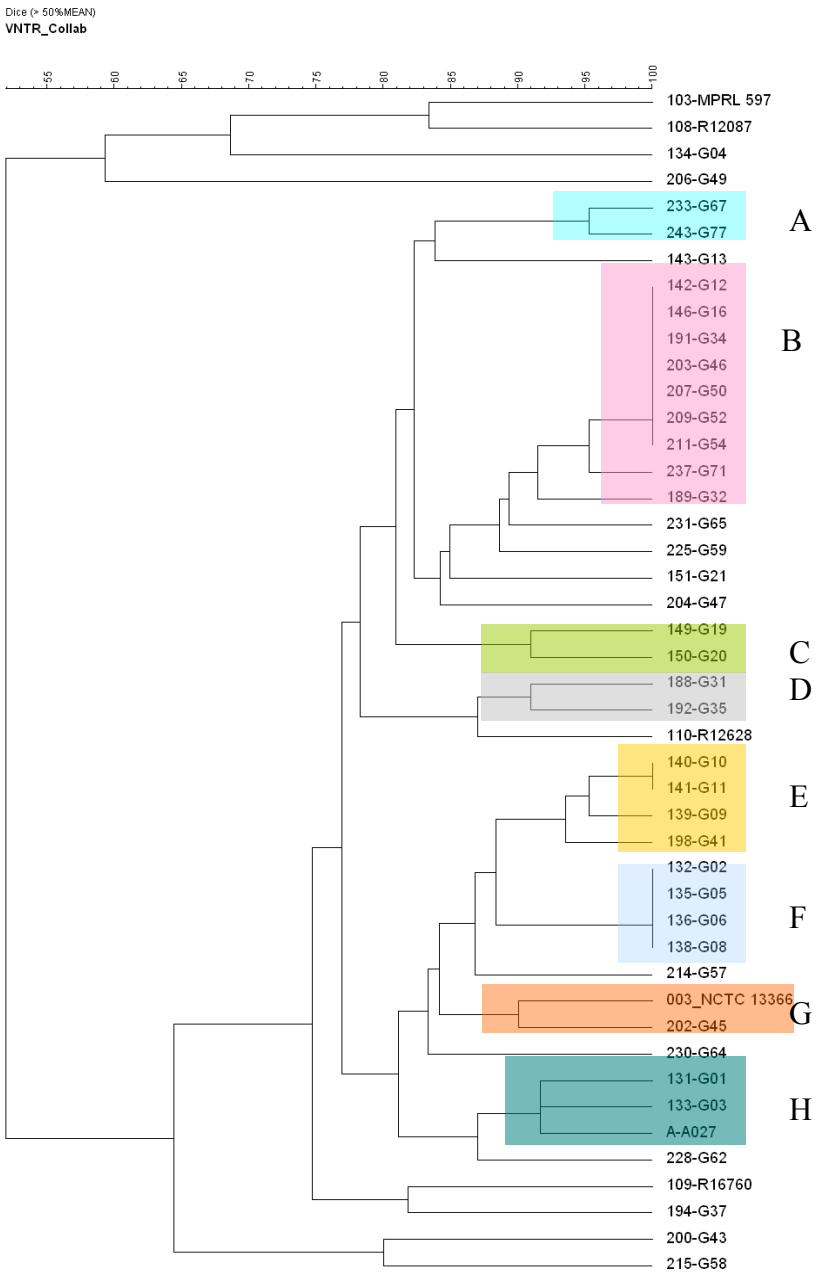


Figure 3.34: Cluster analysis of 45 ribotype 027 isolates, excluding the 14 loci (see Table 3.4) with identical copy numbers for most of the isolates (Table 3.4). Eight groups (highlighted A-H) were detected within ≥ 90 % similarity.

Further analysis was performed on the set of 39 VNTR loci which were identified in this study, excluding the previously published VNTR loci i.e. CD1-10 (van den Berg and Marsh) and the 14 common loci. This was to determine the discriminatory power of the 'new' VNTR loci designed for this study.

Similar to previous cluster analysis, arbitrary value of ≥ 90 % was considered to look at the similarity of the isolates. As a result, seven groups (A-G) were identified among the isolates (Figure 3.35). Group A consisted of 12 isolates with 100 % similarity to each other. Group B with two isolates showing different copy numbers in loci CD20 (null) and CD43 (21) was detected. In group C, two isolates with different copy numbers in three different loci could be observed. Group D contained four isolates and three of these showed 100 % similarity to each other. The remaining isolate differed in copy numbers in loci CD24 (1) and CD46 (6) compared to the others. In group E, 4 isolates showed 100 % similarity to each other while one isolate differed in locus CD12 (null) to the others. Group F contained the recently sequenced reference strain 630, NCTC 13366, which showed 100 % similarity to isolate G41. Another strain isolated from the Stoke Mandeville outbreak (A027), clustered separately and showed 69 % similarity to NCTC 13366. The last group, G, was detected with two isolates differing in loci CD46 (6 and null). Interestingly, isolate G49 clustered separately from the rest of the isolates in both cluster analysis showing a lower similarity percentage, 47 %, to the rest of the ribotype 027 isolates. This isolate was later identified as non-027 isolate from fluorescent ribotyping at HPA, London.

Comparing the two separate cluster analysis performed for ribotype 027 isolates, the ‘new’ primers were able to differentiate most of the ribotype 027 isolates, indicating these loci could be used as markers to differentiate among the ribotype 027 isolates. In particular, loci CD14 and 23 could be used as markers to differentiate ribotype 027 isolates from rest of the ribotypes.

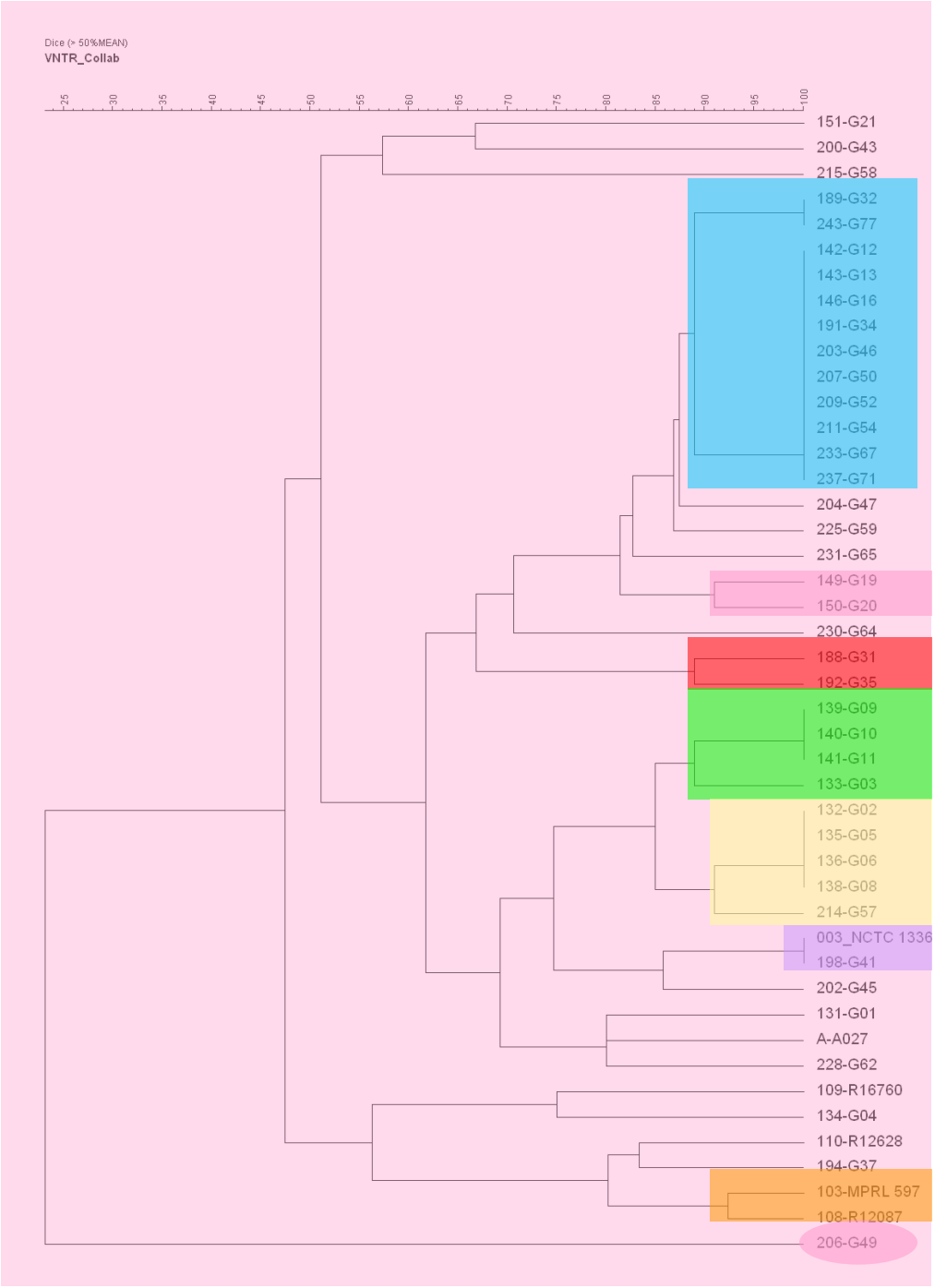


Figure 3.35: Cluster analysis of 45 ribotype 027 isolates, only with the primers designed for this study (i.e. excluding the published loci CD1-10 and 14 loci with identical copy number for all isolates). Seven groups (highlighted A-G) were identified with isolates showing ≥ 90 % similarity. Isolate G49 clustered separately, exhibiting only 47 % similarity to other ribotype 027 isolates and was later identified as a non-027 isolate.

VNTR Locus	Location	Size (bp)	Copy no. range
CD 1	755721-755950	6	14-37
CD 2	3688632-3688751	7	7-27
CD 3	3239736-3239835	6	9-39
CD 4	167124-167172	7	3-10
CD 5	1954913-1954939	3	5-6
CD 6	664660-664705	8	4-16
CD 7 β†	4116072-4116109	9	2
CD 8	692929-693015	8	3-16
CD10	677132-677386	17	1-18
CD11 β†	2168250-2168268	9	Null
CD12 β	804519-804537	9	Null*
CD13	301056-301074	3	5*
CD14 β	788448-788465	9	2*
CD15 β	800026-800043	9	2*
CD18	3014724-3014741	9	1*
CD19	2167938-2167955	9	Null*
CD20 β	2169070-2169086	7	Null*
CD21 β	2169418-2169435	9	Null*
CD22 β†	804623-804645	3	7
CD23 β†	804652-804699	18	1
CD24 β	478055-478070	7	2*
CD25	543601-543619	9	Null*
CD 27	3246180-3246195	6	2*
CD 28	3246752-3246770	9	Null-2
CD 30 β†	794475-794493	7	2
CD 32 β†	252622-252644	6	3
CD 33 β†	800719-800741	2	10
CD 34 β†	799700-799717	6	6
CD 35 β†	797131-797146	7	1
CD 36 β†	881618-881637	10	2
CD 37 β†	881702-881718	8	1
CD 38 β†	883903-883918	6	2
CD 39 β	884179-884194	6	2*
CD 40 β	884601-884612	3	4*
CD 41 β	886202-886220	9	2*
CD 42 β	923152-923190	20	19*
CD 43	623650-623691	21	10-21
CD 45	3319803-3320164	45	4-6
CD 46 β	3753183-3753574	45	6*

Table 3.5: Summary of the VNTR copy numbers for 25 ribotype 001 isolates for 39 loci. Loci highlighted in red were observed with 'null' results for most of the isolates.

* Indicates majority of the isolates produce the same copy number and only a few isolates exhibit one or two different copy numbers
† Indicates stable loci with identical copy numbers for majority of the isolates.
β Loci (23 loci) excluded from the first cluster analysis.

3.9.5 Cluster analysis of ribotype 001 isolates

Among the 25 ribotype 001 isolates, 12 loci with identical copy numbers were observed within the 39 loci (Table 3.5). One of these 12 loci exhibited 'null' results for all 25 ribotype 001 isolates (CD11). Also for another six loci, null results were detected for 90 - 100 % of the isolates (Table 3.5). Ten loci (CD11, 12, 14, 18, 19, 20, 21, 22, 23 and 25) exhibited differential copy numbers for ribotype 001 and ribotype 027 isolates (Table 3.4 and 3.5), Additionally for a further 11 loci, most of the isolates were detected with an identical copy number while one or two isolates either showed a different copy number with one or two repeats difference or a null result (Table 3.5).

VNTR loci CD1-8 and 10 exhibited diverse copy numbers within ribotype 001 isolates and these loci were earlier published by Marsh and van den Berg (Table 3.5). According to the study by Van den Berg, for loci CD4, 5, 6 and 7 copy numbers of 5-7, 5, 6-8 and 2 was reported respectively. However in this study, only locus CD7 showed an identical copy number of 2 for all the ribotype 001 isolates and locus CD5 isolates had copy numbers of 5-6.

When performing cluster analysis, 23 loci that exhibited identical copy numbers for all the isolates or loci that showed variable copy numbers in only one or two isolates (Table 3.5 and Figure 3.36), were excluded from the analysis. Similarly, an arbitrary value of $\geq$ 90 % similarity detected three clusters (A-C) among the 25 isolates. Group A consisted of two isolates exhibiting variable copy numbers within published loci CD1-10. Group B consisted of four isolates; of these three showed 100 % similarity to each other while one isolate, MPRL 587, showed a 92 % similarity to the other three. This isolate varied in copy numbers to the rest in loci CD18 (copy number-1) and CD43 (copy number-21). The last group C contained four isolates showing 100 % similarity to each other.

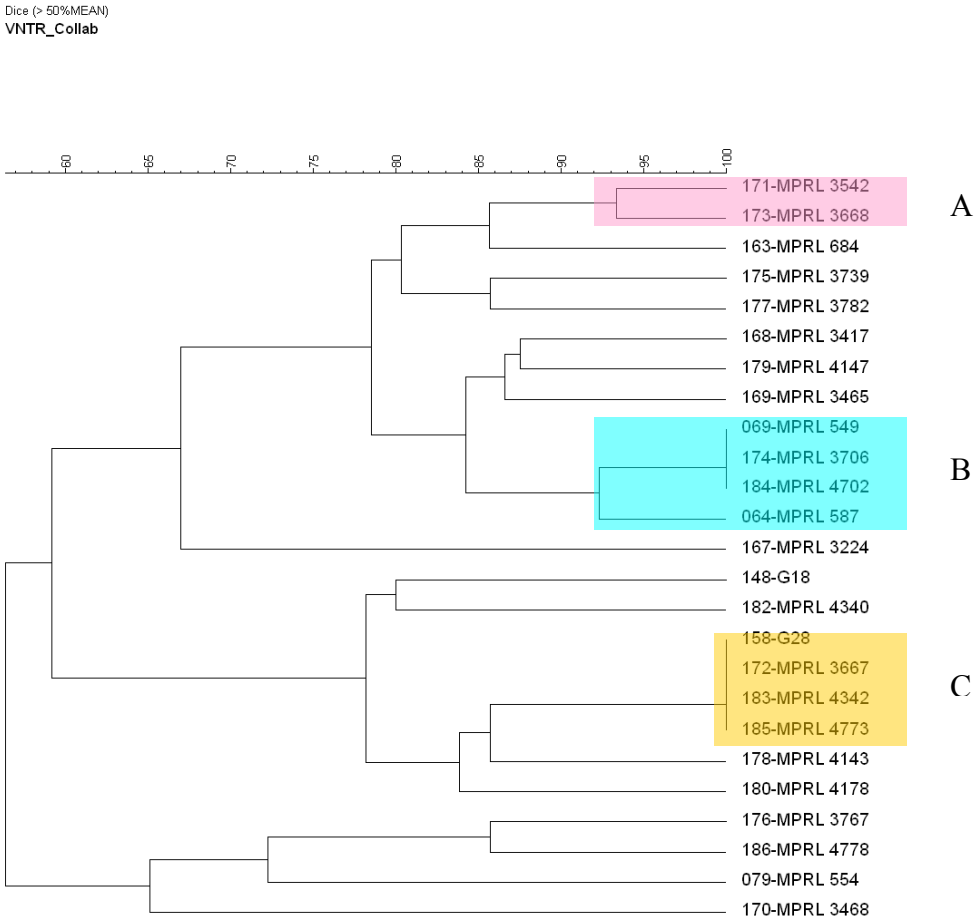


Figure 3.36: Cluster analysis of 25 ribotype 001 isolates, excluding the 23 loci with identical copy numbers for most of the isolates (Table 3.5). Three groups (highlighted A-C) were detected with ≥ 90 % similarity.

Cluster analysis was repeated with only the 'new' loci from the previous cluster analysis in order to detect the discriminatory power of the 'new' loci (Figure 3.37). An arbitrary cut-off of ≥ 90 % was set to look at the similarities among the isolates and four groups were detected (A-D). Group A consisted of three isolates; two of these isolates showed 100 % similarity to each other and one isolate showed a similarity of 90 % to the other two. This isolate (MPRL 684) showed a different copy number for loci CD18 (null) and CD45 (5) while the other two showed copy numbers one and four respectively. Groups B and D contained 4 and 7 isolates respectively and showed 100 % similarity to each other within the group. Group C consisted of two isolates and both showed different copy numbers in loci CD18 (null-1) and CD43 (20-21) and exhibited 90 % similarity to each other.

From the two separate cluster analysis performed for ribotype 001 isolates, the 'new' primers were able to differentiate most of the ribotype 001 isolates. This indicates that these loci could be used as markers to differentiate among the ribotype 001 isolates.

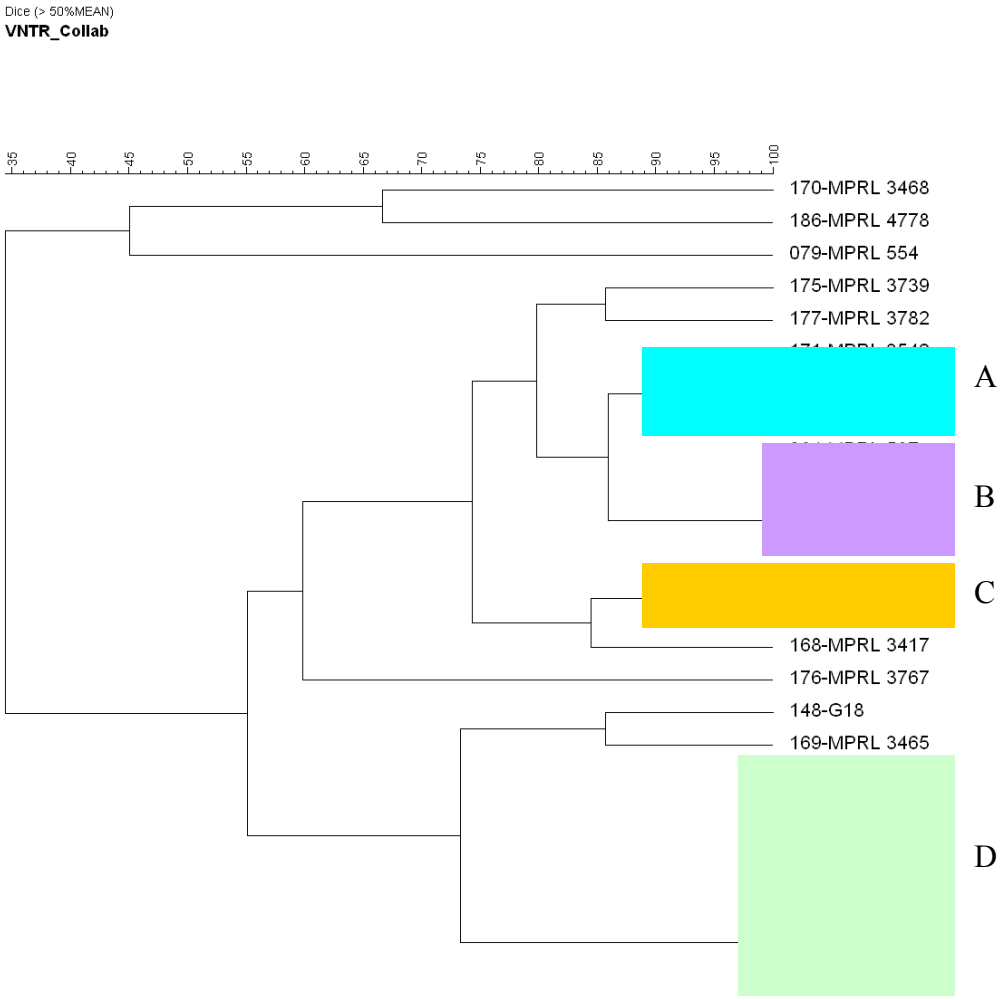


Figure 3.37: Cluster analysis of ribotype 001 isolates with the VNTR loci designed for this study (i.e. excluding loci CD1-10 and loci with identical copy number for all isolates). At the arbitrary cut-off level of ≥ 90 % similarity, four groups (highlighted A-D) were detected among the 25 isolates.

VNTR Locus	Location	Size (bp)	Copy no. range
CD 1	755721-755950	6	1-47
CD 2	3688632-3688751	7	8-28
CD 3	3239736-3239835	6	18-47
CD 4	167124-167172	7	2-10
CD 5	1954913-1954939	3	4-8
CD 6	664660-664705	8	1-15
CD 7	4116072-4116109	9	1-4
CD 8	692929-693015	8	1-28
CD10	677132-677386	17	2-24
CD11	2168250-2168268	9	Null -1
CD12	804519-804537	9	Null -2
CD13	301056-301074	3	Null -5
CD14	788448-788465	9	Null -2
CD15	800026-800043	9	Null - 2
CD18	3014724-3014741	9	1*
CD19	2167938-2167955	9	Null -2
CD20	2169070-2169086	7	Null -1-2
CD21	2169418-2169435	9	Null -2
CD22	804623-804645	3	7*
CD23	804652-804699	18	1*
CD24$^{\beta}$	478055-478070	7	2
CD25	543601-543619	9	Null -2
CD 27	3246180-3246195	6	Null -1-2
CD 28	3246752-3246770	9	Null -2
CD 30	794475-794493	7	Null -2
CD 32$^{\beta}$	252622-252644	6	3
CD 33	800719-800741	2	10*
CD 34	799700-799717	6	6*
CD 35$^{\beta}$	797131-797146	7	1*
CD 36$^{\beta}$	881618-881637	10	2
CD 37$^{\beta}$	881702-881718	8	1
CD 38$^{\beta}$	883903-883918	6	2
CD 39	884179-884194	6	2*
CD 40	884601-884612	3	4*
CD 41$^{\beta}$	886202-886220	9	2*
CD 42	923152-923190	20	9-19*
CD 43	623650-623691	21	10-21*
CD 45	3319803-3320164	45	0-5*
CD 46	3753183-3753574	45	5-7*

Table 3.6: Summary of the VNTR repeats for 22 different ribotype isolates for 39 loci.
Loci highlighted in red showed copy numbers of 1, 7 and 1 for majority of the isolates compared to ribotype 027 and 001 respectively.

* Indicates, majority of the isolates produce the same copy number and only a few isolates exhibit one or two different copy numbers.
β Loci (7 loci) excluded from the first cluster analysis.

3.9.6 Cluster analysis of non-ribotype 027 and 001 isolates

Among the 22 non-ribotype 027 and 001 isolates, variable copy numbers were detected for majority of the 39 loci, Null results could be seen for nearly 90 % of the isolates in 12 different loci (Table 3.6). Stable copy numbers were detected for 6 loci (CD24, 32, 36, 37, 38 and 41) with one locus, CD41 showing a variable copy number for only one isolate (Table 3.5). In loci CD18, 22 and 23, copy numbers of one, seven and one were detected for majority of the isolates respectively and these loci could be potential markers to differentiate from the ribotype 027 and 001 isolates. Loci CD1-10 showed variable copy numbers for most non-ribotype 027 and 001 isolates.

Cluster analysis was performed for the 22 non-ribotype 027 and 001 isolates by excluding seven loci with the same copy numbers or loci, which showed variable copy numbers in only one or two isolates (Figure 3.38). An arbitrary value of ≥ 90 % similarity was considered to detect clusters among the isolates. The different ribotype isolates clustered separately from each other showing high variability within the isolates. A single cluster was detected with five isolates. Two isolates showed 100 % similarity to each other and these two isolates belonged to the same ribotype i.e. 012. The other three isolates showed different copy numbers and showed different similarities to each other (Figure 3.38). From these three isolates, one of the isolate (MPRL 557) belonged to ribotype 002 and the other two ribotypes are unknown.

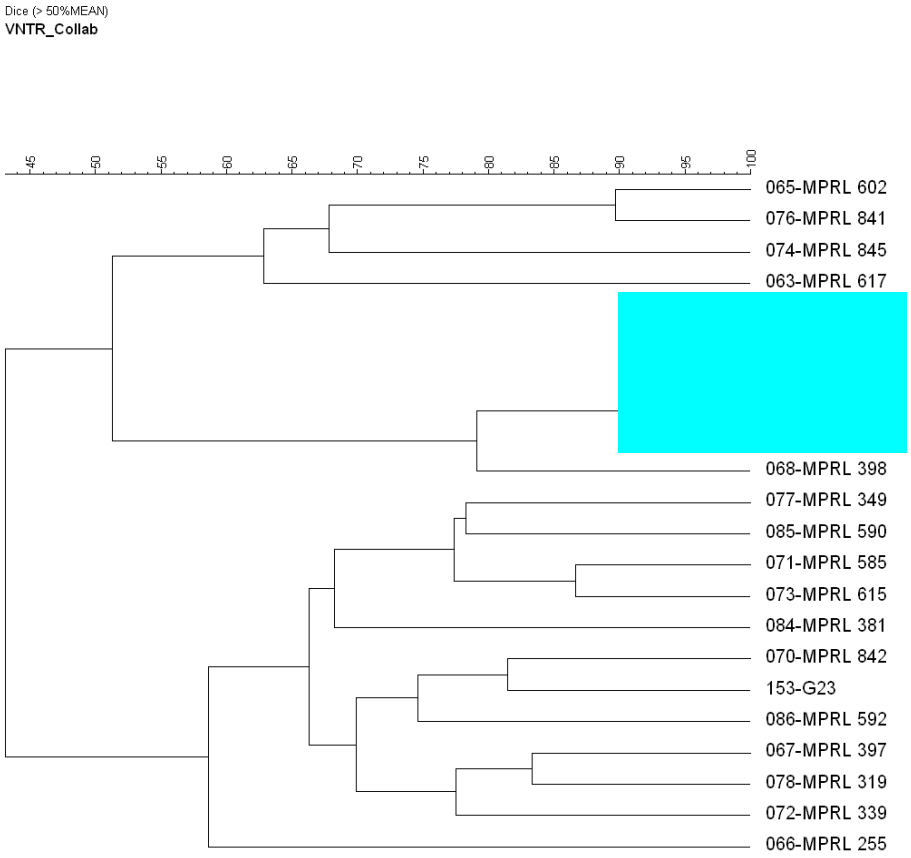


Figure 3.38: Cluster analysis of 22 different ribotype isolates, excluding the 7 loci with identical copy numbers for majority of the isolates (Table 3.6). One group (highlighted) was detected with ≥ 90 % similarity.

Cluster analysis was repeated by excluding the published loci and seven loci with identical copy numbers for most of the isolates, in order to detect if the loci identified for this study were able to differentiate the isolates. Within the ≥ 90 % similarity range, two different clusters were detected (A and B) (Figure 3.39). Cluster A consisted of two isolates MPRL 602 and MPRL 841, showing 91 % similarity to each other. Cluster B was detected with 5 isolates; of these three isolates showed 100 % similarity to each other. Interestingly, out of the three isolates, two were ribotype 012 and the other isolate was ribotype 002. The other two isolates, MPRL 202 and MPRL 595 showed 93 % and 90 % similarities with the other 3 isolates respectively.

Similar to other ribotypes isolates, the 'new' primers were able to differentiate most of the non-ribotype 027 and 001 isolates.

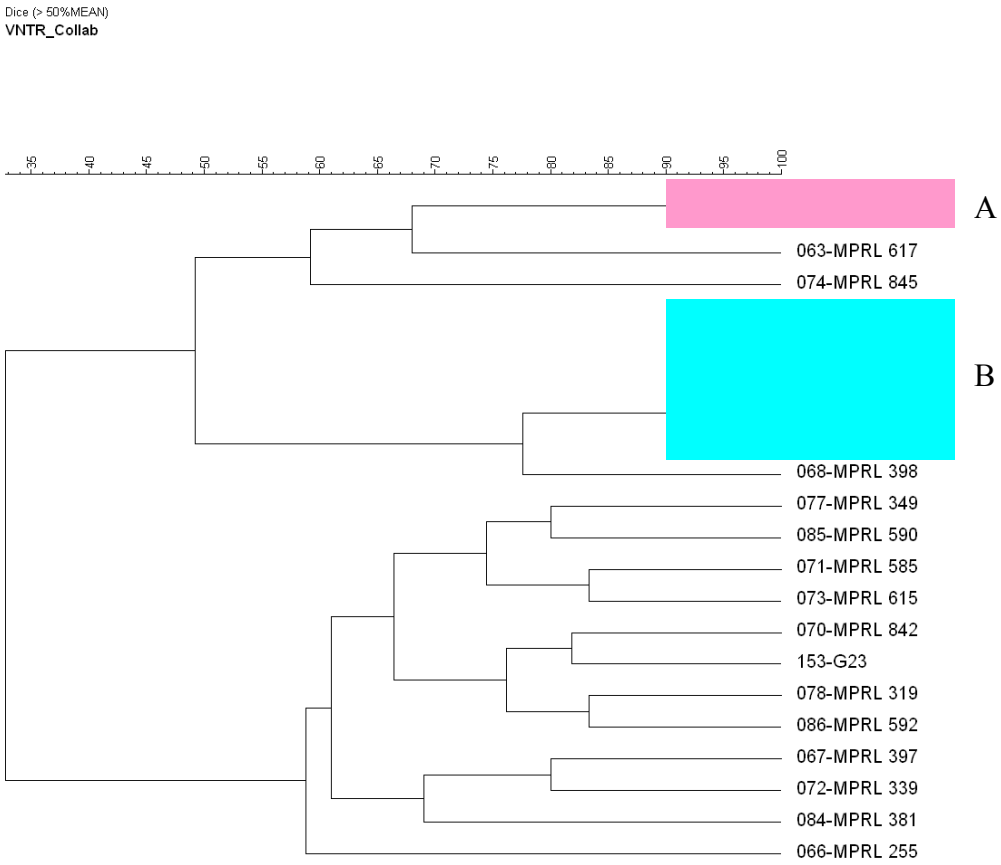


Figure 3.39: Cluster analysis of 22 different ribotype isolate, only with the primers designed for this study (i.e. excluding the published loci and loci with identical copy number for all isolates). Two groups (highlighted A and B) were detected showing ≥ 90 % similarity to each other.

Overall comparison of all ribotype groups and loci

Cluster analysis results from the three ribotype groups were closely compared in order to identify specific loci to differentiate the three groups. Hence, the discriminatory power of the newly designed primers and the published loci i.e. CD1-10 were also assessed.

Cluster analysis was performed again for all ribotype isolates by using the new loci in order to detect the discriminatory power of these loci. Interestingly, two main clusters could be detected and majority of the isolates from the three groups clustered within these two clusters separately from each other (Figure 3.40).

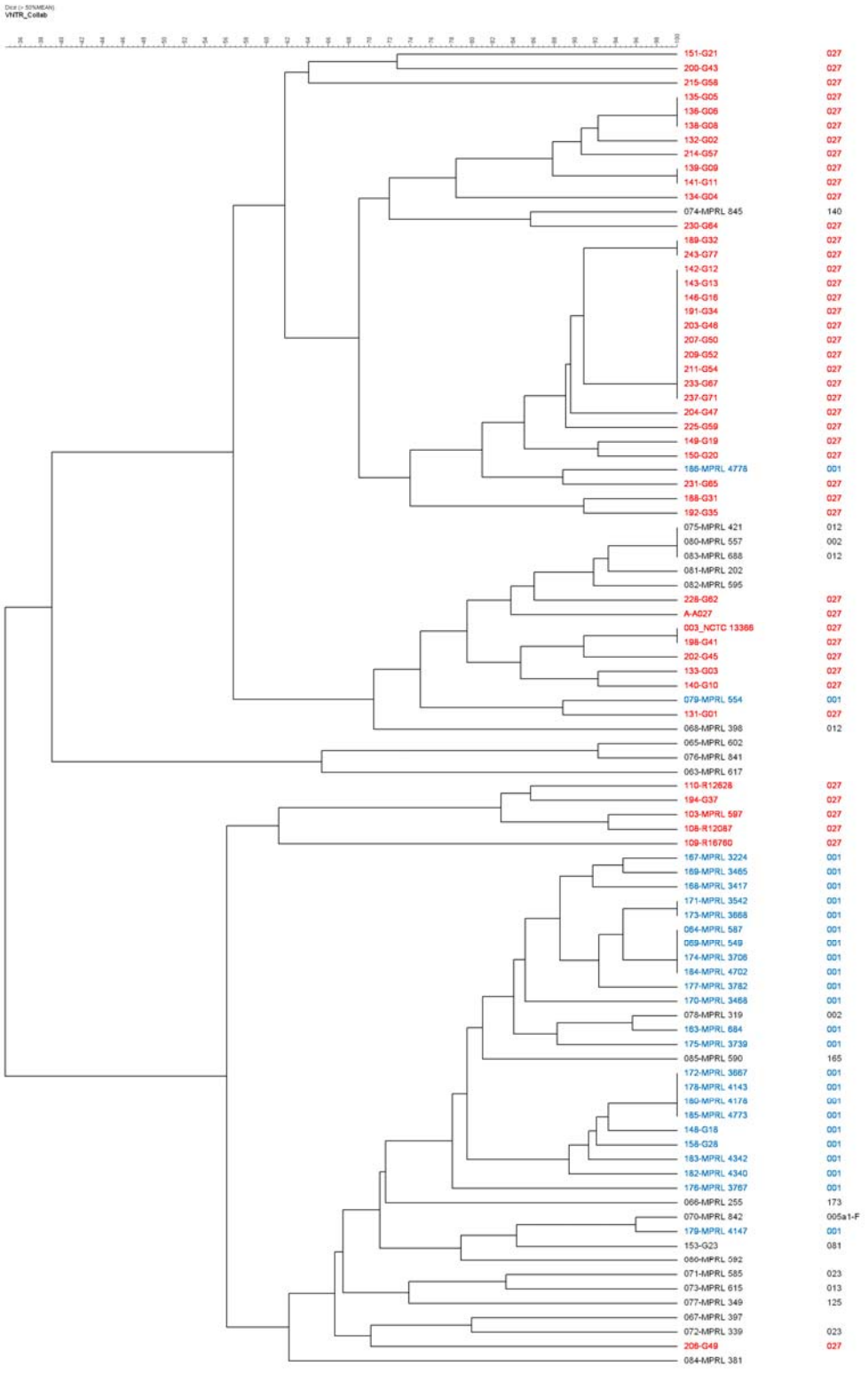


Figure 3.40: Cluster analysis of all three ribotype groups (92 isolates) using the new loci. Majority of the isolates clustered within the particular group. Red: ribotype 027, Blue: ribotype 001, Black: other ribotypes.

Further, analysis of results from the previous clusters and careful observation of the copy numbers obtained for each locus by majority of isolates in each ribotype group revealed that loci CD11, 12, 14, 18, 19, 20, 21, 22, 23 and 25 showed different copy numbers in each ribotype group (Table 3.7). Hence, these ten loci could be used as possible markers to differentiate the 027 and 001 ribotype groups from other ribotype groups.

Loci	Ribotype 027 copy number	Ribotype 001 copy number	Other Ribotypes copy number
CD11	1	Null	1
CD12	2	Null	2
CD14	Null	2	2
CD18	2	1	1
CD19	2	Null	2
CD20	2	Null	2
CD21	2	Null	2
CD22	1	7	7
CD23	0	1	1
CD25	2	Null	2

Table 3.7: Different copy numbers detected within majority of ribotype 027, 001 and other ribotype isolates in new loci.

Using these 10 loci, the cluster analysis was performed again in order to detect if these specific loci were able to differentiate the three ribotype groups (Figure 3.41). Interestingly, again majority of the isolates from the three ribotype groups were clustered separately from each other.

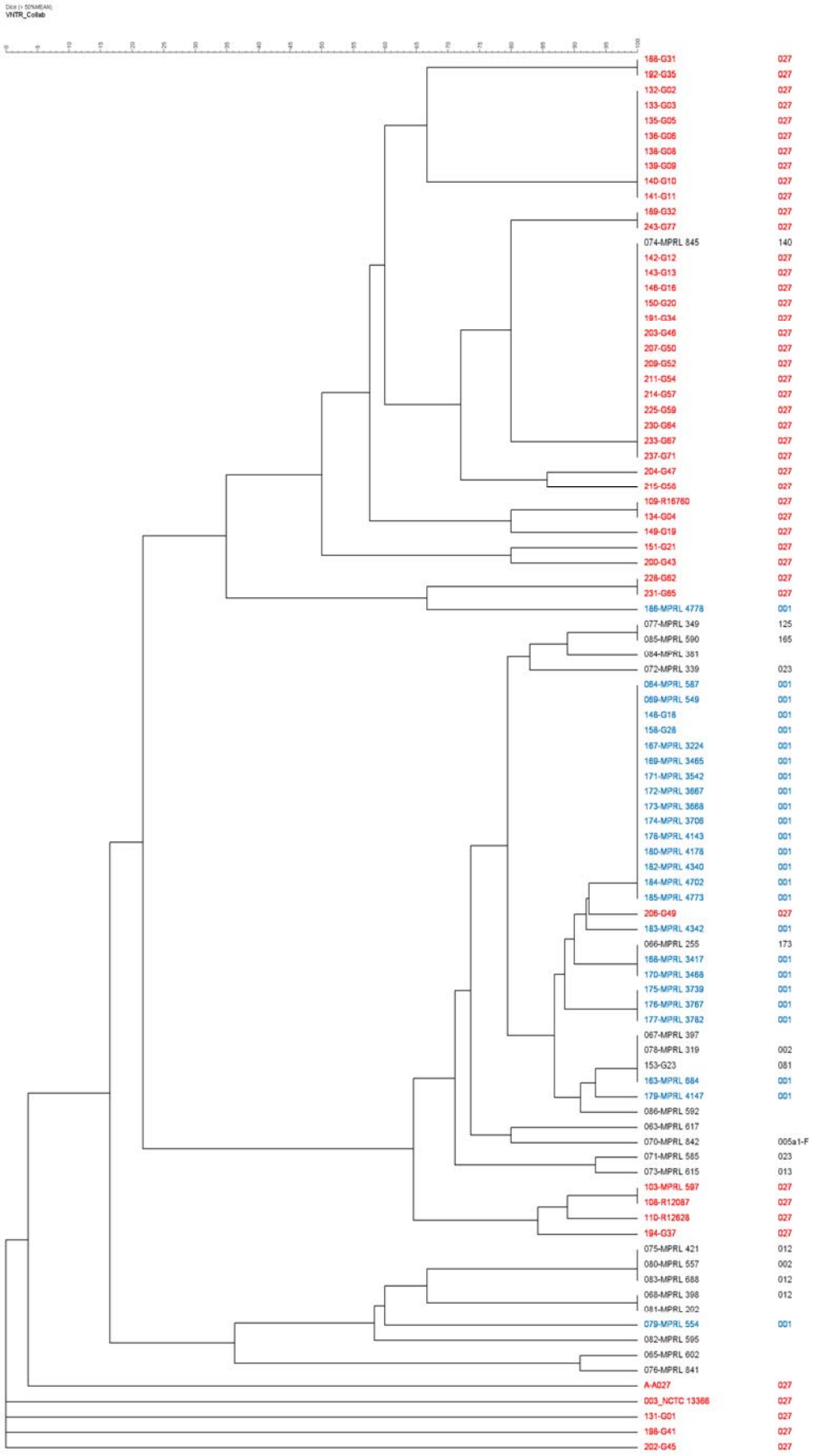


Figure 3.41: Cluster analysis of all three ribotype groups (92 isolates) using the new 10 specific loci. Majority of the isolates clustered within the particular groups. Red: ribotype 027, Blue: ribotype 001, Black: other ribotypes.

Chapter 4

Discussion

4. Discussion

Accurate and definitive microorganism identification and detection is vital for disease diagnosis, treatment, infection control and outbreak investigation. Traditionally, bacterial identification has been based upon phenotypic characteristics, Grams stain, culture morphology, biochemical tests and more recently long chain fatty acid analysis. The importance of these tests is that it is able to yield more direct functional information of metabolic activities taking place in the cells for development, growth and survival. At present, most routine laboratories mainly rely on phenotypic based methods for the identification of bacterial pathogens (Iorio *et al.,* 2007). However, because of the limitations of these methods, genotypic techniques have also been introduced over the years to characterise bacteria. With the wide spread use of PCR, 16S rRNA gene sequencing has become the most commonly used, accurate and reliable molecular based method for the identification of clinically important microorganisms. Also, it is one of the main tools used to detect taxonomic relationships among the bacteria (Clarridge 2004; Bergey's manual 2009). However, as in other methods, it has its limitations and drawbacks (Woo *et al.,* 2008).

Over the years, with developments in the field of proteomics, mass spectrometry has become one of the major tools in microbiology due to its high sensitivity, detection limits and numerous applications (Graves and Haystead 2002; Cañas *et al.,* 2006). Even to date, cellular fatty acid profiling by gas chromatography MS (GC-MS), which was introduced in early days of MS, remains an important tool for taxonomy and classification (Fox 2006). However, with the development of soft ionisation techniques, MALDI and ESI in 1980s by Karas and Hillenkamp and Fenn respectively led to the analysis of large molecules such as proteins comprising several hundred thousand daltons (Tanaka *et al.,* 1988; Karas 1988). The main differences in terms of methods were that, the sample can be directly spotted to a target plate for MALDI analysis while with ESI; samples are sprayed in to the MS. One of the main advantages of MALDI over ESI is that it is more tolerant to sample contaminants like buffers and salts when they are present in low concentrations (Cañas *et al.,* 2006).

In the field of microbial identification, MALDI has become more popular due to the direct analysis of intact cells which made the method much simpler, thus reducing the sample preparation time, as compared to other analytical methods (Claydon *et al.,* 1996; Edwards-Jones *et al.,* 2000; Jarman *et al.,* 1999). This method enabled the detection of mostly

surface-associated proteins which are of major importance, as the surface of a cell is the main site of host-pathogen interactions. The MS fingerprints generated from these analyses were shown to be reproducible and specific for several bacterial species (Claydon *et al.,* 1996; Shah *et al.,* 2000). Hence this method began developing as a fingerprinting tool to characterise biomarkers from microorganisms by assembling databases of mass spectral fingerprints containing reference and type strains (Keys *et al.,* 2004). This enabled rapid high-throughput detection of microorganisms and the potential for use, especially in routine diagnostic laboratories was evident. To date a major problem faced in routine diagnostic labs is the lack of rapid techniques to identify microorganisms in time to influence patient treatment. Although there are many rapid automated or semi-automated diagnostic systems available for bacterial identification, each method has its limitations. Hence the aim of this study was to use MS and other proteomic techniques to characterise the nosocomial pathogens, *Staphylococcus aureus* and *Clostridium difficile*.

With the rising number of MRSA and *C. difficile* associated infections globally due to rapidly evolving strains, the need to detect these organisms at the earliest stages of infection is even more important to prevent further spread. To date the identification of *S. aureus* is routinely achieved using phenotypically based techniques (Kloos and Wolfshohl 1982; Iorio *et al.,* 2007 and Delmas *et al.,* 2008). More recently, genotypic methods such as *spa* typing (Swenson *et al.,* 2001) for the detection of the *mec*A gene has become the most popular method for the detection and typing of *S. aureus* (MRSA).

In the case of *C. difficile*, it is mainly detected in the clinical laboratories by culture and by toxin detection. The cycloserine-cefoxitin-fructose agar, a selective media introduced by Lance George (George *et al.,* 1979; Delmee 2001; O'Connor *et al.,* 2001) in the 1970s is still used for presumptive identification of *C. difficile*. Immunoassay (EIA) has been used widely for the detection of toxin A and B as a screening test for *C. difficile*-associated diarrhoea (CDAD) while tissue culture which detects toxin A and B remains the gold standard for the detection of pathogenic *C. difficile* due to its sensitivity and specificity (Delmee 2001; Snell *et al.,* 2004; Russmann *et al.,* 2007). There are also various commercial kits for the detection of CDAD with different sensitivities and specificities (Reyes *et al.,* 2007 and Brazier 1998). An additional detection method which involves the use of the common antibody to the enzyme, GDH present in all strains of *C. difficile,* is used by some laboratories (Wilkins and Lyerly 2003).

With the increasing number of samples which need to be identified correctly, rapid, high-throughput techniques have become necessary in diagnostic laboratories. Such laboratories are now turning towards the emerging technologies to resolve these issues.

MALDI-TOF-MS analysis of *S. aureus* and *C. difficile*

This part of the study explored the use MALDI-TOF-MS as a tool for the characterisation of bacteria. The performance of an existing database (containing profiles of more than 5000 isolates) and a search engine to match the spectra of clinical isolates to the profiles of type strains and laboratory reference strains was investigated. Furthermore, the relative similarity of the MALDI fingerprints amongst all *S. aureus* isolates and other closely related *Staphylococcus* spp were also tested. For the study, 95 clinical isolates presumptively identified as *S. aureus* collected from The Royal London Hospital were confirmed as *S. aureus* using conventional biochemical tests by the Staphylococcus Reference Laboratory at HPA.

The profiles generated by MALDI were mainly from surface-associated molecules, which in turn could be affected by environmental pressure. Hence, parameters such as variation in culture media, pH, growth time and incubation temperature were re-examined briefly, although these were standardised prior to assembling the database (Shah *et al.,* 2000, 2002). Although the MS of the MSA medium showed greater peak density compared to CBA within the mass range of 500- 3000 Da, CBA was chosen for the study because the mass ion density was still significant for cells grown on this medium. This is in agreement with the study of Walker *et al.* (2002) who showed that for *S. aureus* the reproducibility and a higher number of mass spectral ions could be obtained using CBA. Thus, CBA was also used as the standard medium for assembling the database (Figure 3.1). Similarly, the spectra from different incubation periods showed that with increasing incubation time, the quality of the spectra decreased. Since the spectra from 24 h incubation times showed the largest density of peaks, the study was carried out using cells grown on CBA for 24 h; a factor that will be also beneficial for clinical laboratories due to the reduced turn-around times.

The first phase of the study revealed that with the exception of three isolates (HPA 80, HPA 547 and HPA 549), all 92 isolates matched to spectra in the database and were correctly identified by the software. However, almost all of the 95 isolates showed similar

profiles with several consensus mass ions in the range of ~ 800 - 3500 Da and could be regarded as specific surface biomarkers for the genus *Staphylococcus*. Within the same mass range, Edward-Jones *et al.* (2000) and Du *et al.* (2002) also reported the presence of specific biomarkers for *S. aureus* and MRSA. Subsequently, Bernado *et al.* (2002) also reported the presence of specific markers in this mass range although that study was not based on intact bacterial cells and therefore highlights the reproducibility of the methodology. For some of the clinical isolates, unusual peaks were detected in the mass range of ~2000 - 3400 Da which were absent from reference/type strains indicating possible phenotypic differences. From the parallel study done with the 2006 database at MMU to validate the methodology, contamination of these samples were again highlighted by their incorrect identification, RMS and probability values. Consequently these were purified by sub-culture, re-run and the results were concordant with other isolates (Table 3.1). The presence of occasional contaminants in a sample that leads to altered RMS and probability values emphasises the high sensitivity of the technology which allowed for detection of mass ions from the contaminants even before they were visually detected on the agar plates and suggest that in the future the analysis of mixed samples may be possible.

The second phase of the study contained 39 clinical isolates from a collection of temporally, geographically and genotypically diverse strains of *S. aureus* including MSSAs along with representatives of epidemic HA-MRSA (EMRSA-15, -16 and -17), CA-MRSA (so-called USA300, South West pacific and European clones) and two 'aberrant' strains (one coagulase-negative *S. aureus* and one small colony variant). Their successful identification to the species level by MALDI-TOF-MS gives further support to this approach. However despite good representation of MRSA strains analysed in the present study (~ 44 %), there were no consistent mass ions that distinguished between MRSA strains from other *S. aureus* strains that were reported by Edwards-Jones *et al.* (2000) and must cast doubt on the validity of this approach for antibiotic resistant profiling.

Overall the results of this part of the study demonstrated clearly that isolates of *S. aureus* may be identified by MALDI-TOF-MS with a high degree of confidence (97.76 %) in a clinical laboratory using an established database. However, reliance on a few distinguishing biomarker peaks to differentiate populations of *S. aureus* into biological types nearly always begin deviating as more wild-type isolates are studied, therefore the entire mass spectrum should be taken into consideration when devising a classification

scheme. In a similar attempt to identify coagulase-negative staphylococci (CNS) using a database containing 23 reference strains, Carbonelle *et al.* (2007) showed clear separation of coagulase-negative staphylococci from other closely related taxa. A recent study by Dupont *et al.* (2009) also demonstrated the robustness and high sensitivity of microbial identification using the same database used by Carbonelle for his study, by MALDI technology for clinical CNS isolates. But in his study, he compared MALDI with two automated systems, BD Phoenix and VITEK-2, phenotype based techniques for comparison. Excluding the species that were not present in the database, they were able to achieve a correct identification for all isolates with a percentage of 97.4 %, 79 % and 78.6 % for MALDI, Phoenix and VITEK-2 system respectively. The previous studies carried out for the identification of *S. aureus* based on phenotypic properties using various commercial kits, such as API Staph-Ident kit, DMS Staph-Trac kit, MicroScan Pos ID and commercial slide agglutination tests showed variable results. Kloos and Wolfshohl (1982) was able to detect a 90 % similarity between conventional methods and the API Staph-Ident kit while Overman and Overley (1990) reported between 75 – 100 % reproducibility of results using two versions of the Staph-Ident kit. By using DMS Staph-Trac kit, Giger *et al.* (1984) achieved 88 % correct identification for *S. aureus*. A recent study by Layer *et al.* (2006) compared three methods; BD Phoenix, VITEK 2 and API ID32 STAPH test, for the detection of *Staphylococcal* species from clinical samples. They demonstrated that for the identification of CNS in routine laboratories, both BD Phoenix and VITEK 2 systems are within highly acceptable limits, although API ID32 STAPH revealed more correct results compared to both automated systems.

Several molecular methods have also been developed to distinguish *S. aureus* isolates. A study carried by Peacock *et al.* (2002) for *S. aureus* isolates in a renal unit demonstrated similar discriminatory capabilities of both MLST and PFGE typing methods. They concluded that due to the time, cost, and expertise required to perform MLST, PFGE will remain the typing system of choice for outbreak investigations (Weller 2000), with MLST being reserved for long-term epidemiological and population genetic studies. A similar study carried out by Melles *et al.* (2007) compared three methods, MLST, PFGE and AFLP, in order to type *S. aureus* isolates. They demonstrated that although there were discrepancies between the results of the three typing methods, all three showed similar results. They also highlighted the drawbacks of these methods such as the lack of reproducibility of the fragment patterns and the difficulty of the data exchange between the laboratories.

The present study assessed the potential of MALDI-TOF-MS of intact *S. aureus* cells as a diagnostic tool in the clinical laboratory and evaluated the performance of an existing database developed for microbial identification (Keys *et al.,* 2004). The method described over the years and used here (Keys *et al.,* 2004; Shah *et al.,* 2008) was aimed at rapid characterisation of microorganisms to the species level and was intentionally devised to facilitate application in a high throughput clinical laboratory. To date, MALDI-TOF-MS of intact cells has been used widely to characterise bacteria. Such an application necessitates development of a microbial database in which MS spectra are attained under highly controlled experimental parameters using type and reference strains from an accredited culture collection such as NCTC. The database used in the present study currently holds MALDI-TOF MS profiles of more than 5000 bacterial strains. Here, the performance of this database and the MicrobeLynx™ search engine to match the spectra of clinical isolates (which may differ) to the profiles of NCTC strains was tested for the first time.

A major aim of this study was to collect field isolates of *S. aureus* from a clinical laboratory and, with minimal culturing time, to adapt to the laboratory conditions, challenge the database and the search engine to match their mass spectral profiles to reference spectra of *S. aureus*. All but four of the 134 clinical isolates presumptively identified as *S. aureus* were correctly identified using MALDI-TOF-MS. Four aberrant isolates gave discrepant results, but were later confirmed as *S. aureus* by 16S rRNA analysis. Notably, they were shown to have low level contamination which clearly affected the quality of their mass spectra.

Spectral patterns of closely related species of *S. aureus* such as *S. epidermidis, S. haemolyticus* and *S. saprophyticus* showed a high degree of similarity across the genus and there were several genus-specific biomarkers. It was possible to ascertain the intraspecies phenotypic diversity among clinical isolates of *S. aureus* and their similarity to type/reference strains by comparative analysis of MALDI spectra. However, the overall similarity in the profiles (as illustrated by the RMS value) rather than specific mass ions were used for identification of strains. Despite the observed differences, there were enough stable characteristics in the profiles of *S. aureus* to achieve correct identification of isolates (Rajakaruna *et al.,* 2009).

A similar approach was carried out for the identification of *C. difficile* using the existing database. However, the three *C. difficile* isolates (L, B1 and T; see Material and Methods) tested using the CBA culture media, were unable to achieve correct identification using the MicrobeLynx™ software. This may have been due to the limited number of *C. difficile* spectra present in the database. When the growth medium was changed from CBA to FAA, the identification improved but only to the genus level in a few cases. This might be because the FAA medium was more suitable than CBA for the growth and the isolates may have expressed more characteristic mass ions on FAA. When comparing the profiles of the predicted isolates by the software (Figure 3.5), similar mass ions were seen for all the entries (1450 Da and 2600 Da) indicating that those predicted species maybe closely related including *C. difficile*. This may have been the reason the software was unable to differentiate between the three *C. difficile* isolates. Hence, in an attempt to identify *C. difficile* using MALDI, another database (SARAMIS™) was used for further analysis.

Spectral ARchive And Microbial Identifications System (SARAMIS™)

The SARAMIS™ database contains more than 100,000 spectral entries representing more than 1,200 species and 220 genera. Microbial identification is achieved through the so called SuperSpectra™, generated for each taxon i.e. genus, species and strain biomarkers. These specific SuperSpectra™ or biomarkers which are representative of respective groups of microorganisms are then used for automated identification of unknown isolates. Although the SuperSpectra™ tool is used to identify an unknown, a reference database is generated by continuously adding new spectra for each new species. To date more than 2,400 SuperSpectra™ are available for automatic microbial identification. Also with the software installed in MS e.g. AXIMA MALDI-TOF-MS (Shimadzu, UK), it takes less than 10 sec per 500 laser shots and analysis is rapid and convenient.

To obtain a SuperSpectrum™, the raw mass spectra from 15-20 strains of a species are collected from different environments, laboratories grown on different culture media etc., and processed to yield peak lists and clustered to obtain the predominant mass ions of a specific species (Figure 4.1). This consensus spectrum is calculated for the mass signals that have been recorded at the frequency of 50-100 % excluding unspecific signals. Hence the consensus spectra contain peaks that are specific for higher taxonomic units such as genera or families in order to facilitate final identification to the species level. When the

SuperSpectra™ are calculated, each non-specific signal is identified by comparing with the reference database. For species identification, these spectra are given a low weighting. In cases where numerous isolates of a single species are analysed, cluster analysis tends to reveal sub-specific units and multiple SuperSpectra™ are calculated for that particular species.

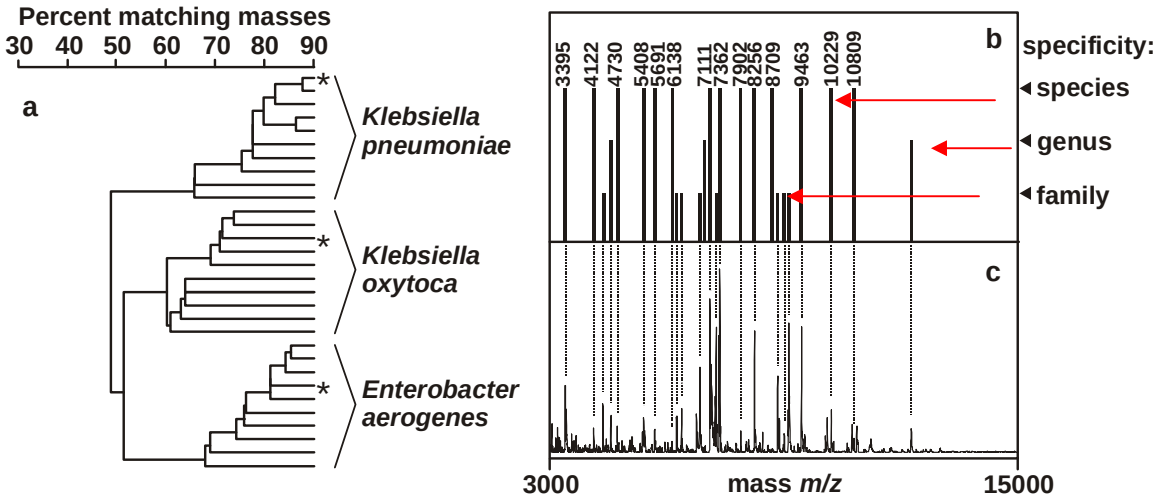


Figure 4.1: Dendrogram of mass spectral fingerprints of ten isolates each of three closely related species within the diverse family Enterobacteriaceae (a), illustration of a SARAMIS SuperSpectrum for *Klebsiella oxytoca* with taxonomic specificities of mass signals indicated by their height (b), and a mass spectrum of a clinical isolate of *K. oxytoca* with mass signals matching to the SuperSpectrum™ indicated by dotted lines (c). Asterisks (*) indicate DSM reference isolates of respective species.

When a mass spectral fingerprint is matched to the SuperSpectra™ in the database, weighting values for each matching signals are summed up. As a result those SuperSpectra™ for which the highest specific concordance was established are listed. These sums of the mass signal values are then translated into a confidence value for identification and also indicated by a colour coded system by the software as shown here for *C. difficile* (Table 4.1).

Colour code	Confidence value
Green	90-100 % (automatic identification)
Yellow	< 80 % (manual identification)
Grey	Spectra not acquired
Red	Mixed culture/poor sample preparation

(a)

4	4F3	MISU 144	MPRL 617	48h	NA		no spectrum	
4	4F4	MISU 144	MPRL 617	48h	NA		no spectrum	
4	4G1	MISU 101	MPRL 002	48h	CBA	96,4%	Clostridium	difficile
4	4G2	MISU 101	MPRL 002	48h	CBA	99,9%	Clostridium	difficile
4	4H1	MISU 102	MPRL 059	48h	CBA	96,4%	Clostridium	difficile
4	4H2	MISU 102	MPRL 059	48h	CBA	99,9%	Clostridium	difficile
4	4H3	MISU 103	MPRL 1037	48h	CBA	91,0%	Clostridium	difficile
4	4H4	MISU 103	MPRL 1037	48h	CBA	81,0%	Clostridium	difficile
4	4I1	MISU 104	MPRL 13366	48h	CBA	90,0%	Propionibacterium	acnes

(b)

Table 4.1: (a) Confidence values of identification by the SARAMIS™ software. (b) Example of an output result by the software for *C. difficile* isolates.

Compared to the previous Waters® method, the main difference of this method was the change of matrix solution from 5-chloro-2-mercaptobenzothiazole (CMBT) and alpha-cyano-4-hydroxy-cinnamic acid (CHCA) to 2, 5-dihydroxybenzoic acid (DHB) in the presence of trifluoro acetic acid (TFA) and acetonitrile (ACN). Several studies have shown that experimental factors such as sample preparation, type of matrix, matrix solvent, instrument quality etc. significantly influence the MALDI profiles and the spectral quality (Williams *et al.,* 2003, Shah *et al.,* 2000, Wang *et al.,* 1998). Hence, the effect of the 'new' matrix solution, DHB, on intact cells of *C. difficile* was investigated (Section 2.3.4). In brief, cells were held at different time intervals in DHB and its ultra-structure was examined by transmission electron microscopy. Disruption of the capsule and external polymeric material around the cells were clearly evident. However, the cell appeared to stay intact on the MALDI target plate (Figure 3.9). This is surprising as TFA and ACN are highly polar solvents and would readily cause lysis of cells in free solution (Williams *et al.,* 2003). A similar observation was also reported by Pennanec *et al.* (2010) for *Legionella*, when the cells were introduced to 2.5 % of TFA. He observed the cells were no longer viable with addition of 2.5 % TFA in the presence of ACN and the observed peaks correspond to membrane and internal proteins.

In the present study, the images obtained from the target plates before and after the MALDI analysis revealed a uniform layer of cells which clumped to leave areas of the target plate clear (Figure 3.10 a). Closer inspection of these areas before analysis showed that the cells were arranged in parallel rows, indicating that the matrix has held the cells in a 'frozen' state. This may have been brought about by the rapid evaporation and crystallisation following addition of the matrix to the cells. The images obtained after the MALDI analysis revealed that the cells are less uniform and showed a metallic sheen indicating the disruption of the surface polymers. A similar observation was reported by Welham *et al.* (1998) for *E. coli* cells prior to MALDI analysis, where the cells remained intact after the addition of the matrix, 2, 4-hydroxyphenylazobenzoic acid (HABA), which contained 0.1 % TFA as a solvent. In their study, images obtained after the MALDI analysis showed cells that were completely intact, indicating the observed ions may arise from the cell wall. This difference may have been due to the two different matrices; DHB and HABA and the concentration of the TFA present in the matrix solvent. Although Welham *et al.* (1998) showed by using a matrix solvent containing 0.1 % TFA did not affect the bacterial cell wall, in the present study, 0.3 % TFA caused disruption of the surface polymers which were visible from the electron micrographs. The spectral profiles obtained using the SARAMIS™ method must therefore arise from intra cellular proteins.

Although the SARAMIS™ method was standardised, parameters such as growth time and incubation times were re-examined briefly for *C. difficile*. The spectra obtained from the three test media: CBA, FAA and NA, for 24- 48 h showed similar reproducible spectra, despite the variation in the composition of the medium. However, the spectral qualities from the 24 h cultures grown on CBA were superior. Hence, this time period was selected for subsequent analyses. In the first phase, 53 isolates were analysed in duplicate and except for two isolates, all were identified as *C. difficile* with 80-100 % confidence values. Significant peaks were detected for most of the *C. difficile* isolates within the mass range of 4200- 4300 Da, 5400 Da, 6300 Da, 6640 Da and 6720 Da. The two isolates incorrectly identified as *S. warnerai* and *Propionibacterium acnes* were due to the presence of mixed cultures, highlighting the sensitivity of the system. In the second phase, 63 isolates were tested and all were correctly identified by the database. Cluster analysis was performed based on the percentage matching identical masses of the spectra and available *C. difficile* profiles in the database. All isolates showed a high degree of similarity and therefore recovered in a single cluster. Also the MS data from the first set of data obtained using 24

h incubation clustered closely while the incorrectly identified isolates were recovered in a cluster distinct from the rest of the isolates.

In addition, the SARAMIS™ database identified 14 closely related clostridial species correctly with a confidence value of more than 65 %. Although the confidence values were lower than the *C. difficile* identification scores for some of these species, the SARAMIS™ identification correlated with that of comparative 16S rRNA sequencing. Cluster analysis of these different species against *C. difficile* consequently resulted in their separation into different phena with < 30 % similarity of mass ions (Figure 3.17). Overall, the results from this study using the SARAMIS™ database showed ~ 98 % correct identification for *C. difficile* isolates. A similar study done by Grosse-Herrenthey *et al.* (2008) used MALDI-TOF-MS for the detection of 31 different *Clostridium* species. The authors reported the successful identification of *Clostridium* species within minutes using MALDI-TOF-MS combined with a database and bioinformatics software tool. The results of this study also correlated with the 16S rRNA sequencing data and traditional diagnostic techniques for all the species. Thus, when compared to large number of methods described earlier for the detection of *C. difficile,* MALDI-TOF-MS provides a rapid and reliable identification with minimal sample preparation and its high accuracy is likely to contribute markedly as a primary diagnostic tool in the near future.

From the first attempt to identify bacteria using intact cells (Claydon *et al.,* 1996; Krishnamurthy and Ross 1996 and Holland *et al.,* 1996), MALDI-MS has rapidly grown over the years as an identification tool. With the introduction of the ICM technique, the need for a mass spectral fingerprint library of bacterial species in order to detect unknown isolates was essential. Standardised experimental parameters were considered a crucial factor in order to achieve comparable reliable identification (Shah *et al.,* 2000, 2002). However, over the years as it evolved as a rapid, inexpensive and accurate method for bacterial identification, the methodology, software and databases have improved so significantly that many clinical laboratories are now employing it as their primary test in hospitals (see review, Emonet *et al.,* submitted for publication, 2010).

To date, apart from the *C. difficile* and *S. aureus* identification, MALDI has been widely used for the detection of other microorganisms. A database similar to AagnosTec developed by Bruker Daltonik GmbH, Germany (BioTyper) has also been used widely for the identification of microorganisms. A recent study by Seng *et al.* (2009) analysed more

than 1000 clinical isolates by MALDI-TOF-MS using Bruker BioTyper database in parallel with conventional phenotypic identification techniques. They reported a 95.4 % correct identification of bacterial isolates, 84.1 % for species and few isolates up to the genus level (11.3 %) by MALDI. In their study, up to 2.8 % isolates were not identified by the MALDI, due to the lack of accurate reference spectra for particular species. Hence, it was pointed out the critical requirement is the development of an accurate database for bacterial identification. Another study by Eigner *et al.* (2009) evaluated the performance of MALDI for the identification of various clinical isolates in a routine microbiology setting. They tested more than 1100 clinical isolates and 108 reference strains and compared it with phenotypic identification systems and 16S rRNA analysis. Overall results for all the isolates by MALDI were 93.5 % and 95.2 % for reference and clinical isolates respectively. They also reported the accuracy of identification for *Enterobacteriaceae*, non-fermenting gram-negative rods, *Staphylococcus*, *Enterococcus* and *Streptococcus* with MALDI as 95.5 %, 79.7 %, 99.5 %, 100 % and 93.7 %, respectively where most of these species are sent for identification in the routine clinical laboratories. A subsequent study by Bizzini *et al.* (2010) also reported MALDI as a fast and reliable technique to use in clinical laboratories, in which they were able to identify more than 1200 isolates to the species level (93.2 %) by MALDI-TOF-MS.

Similarly, MALDI has been used for the detection of many other pathogens e.g. food and water contaminant pathogens, fungal, yeast and many more. Some of these included the identification of *Legionella* species in water samples (Pennanec *et al.,* 2010), bacterial species associated with food contaminants (Mazzeo *et al.,* 2006), classification and identification of *Salmonella* and *Vibrio* species (Dieckmann *et al.,* 2008 and 2009). Also the identification of other pathogens such as *Burkhoderia cepacia* complex (Vanlaere *et al.,* 2008), fungi (Erhard *et al.,* 2008) and yeast identification (Marklein *et al.,* 2009) by MALDI have also been reported recently. Another recent development of MALDI-MS is the identification of bacteria directly from positive blood cultures/bottles in clinical laboratories (La Scola and Raoult 2010; Maier *et al.,* 2010). Currently the identification of microorganisms growing in blood cultures takes up to two days due to the sub-cultivation and biochemical identification. However with introduction of new sample preparation methods described by La Scola and Raoult (2010), Maier *et al.* (2010), MALDI has proved as a rapid and accurate alternative to classical method of identification of blood cultures in clinical laboratories.

SELDI-TOF-MS and SDS-PAGE analysis of *S. aureus* and *C. difficile*

The main goal in proteomics is to resolve, separate, visualise and identify the proteins from complex mixtures. To date, the predominant technology for protein separation and isolation is polyacrylamide gel electrophoresis and it still remains one of the most effective ways to resolve a complex mixture of proteins. For many proteomic applications, SDS-PAGE is the method of choice to resolve protein mixtures and can be used to resolve proteins with molecular masses of 10 -300 kDa (Graves and Haystead 2002). Surface Enhanced Laser Desorption/Ionisation (SELDI) ProteinChip® technology has the ability to generate rapid protein expression profiles from a variety of biological samples and allow the rapid characterisation of the microbes (Schmid *et al.,* 2005). Unlike the inert MALDI target surfaces, the SELDI protein chip surfaces are uniquely designed to retain proteins of interest from complex mixtures according to their specific properties (Seibert *et al.,* 2004). It has been shown that unlike MALDI, SELDI enables visualisation of spectral profiles containing peptides/proteins in the mass range of 10-150 kDa (Shah *et al.,* 2002), thus, it may be used to detect specific biomarkers. However, when compared to SDS-PAGE, the sensitivity of SELDI-MS remains higher as lower molecular weight proteins can be resolved by SELDI-MS than SDS-PAGE (Figure 4.2). Hence, another aim of this study was to identify biomarker proteins or peptides that may be characteristic for MRSA, MSSA and *C. difficile* using SELDI-TOF-MS and SDS-PAGE.

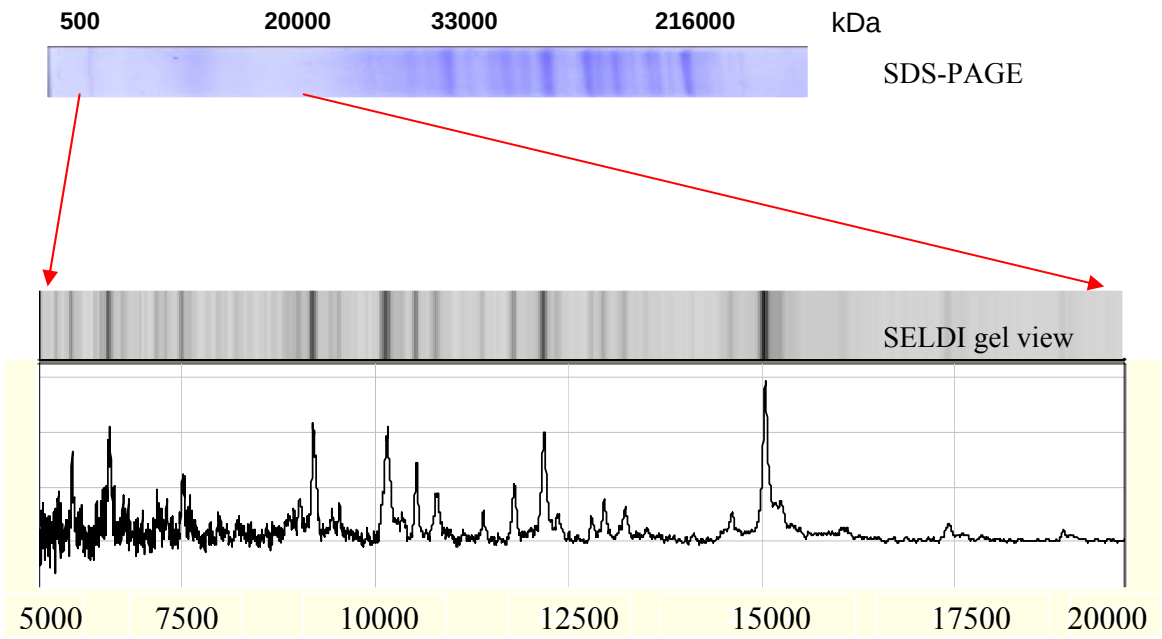


Figure 4.2: Comparison of bacterial cell extracts separated by SDS-PAGE and SELDI-MS. Several peptides/proteins not seen by SDS-PAGE (lower molecular weight) are resolved by SELDI-MS (Taken from Encyclopaedia of Rapid Microbiological Methods. **3**: Figure 1).

For this part of the study, 50 MRSA and 49 MSSA isolates were analysed on different ProteinChip® arrays. Although a previous study (Innes, PhD thesis, 2008**)** suggested that SELDI has the potential to differentiate MRSA and MSSA, a limited number of isolates were tested. In order to reduce the variation of results, some parameters in a method needed to be fully optimised (Cordingley *et al.*, 2003). For this purpose, the same protein extraction method used in previous studies and other methods were tested using different concentrations of lysostaphin, various pH range and different ProteinChips® in order to obtain the optimal cytosolic protein yield for SELDI analysis.

By virtue of its complex cell wall, *S. aureus* is difficult to lyse, however, lysostaphin, an antibacterial enzyme is capable of specifically cleaving the cross-linking pentaglycine bridges of the peptidoglycan of staphylococci (Schindler and Schuhardt 1964). It has also been shown by Wu *et al.* (2003) that lysostaphin not only kills *S. aureus* in biofilms but also disrupts the extracellular matrix of *S. aureus* biofilms *in vitro* on plastic and glass surfaces at concentrations as low as 1 µg/ml. Hence, the use of lysostaphin in this protein extraction method ensures the detection of intracellular cytosolic proteins of *S. aureus* (MRSA/MSSA).

Efficient breakage of cells to release cytosolic proteins is one of the very important steps in proteomic analysis. Here, the total protein content was compared by two different extraction methods: the French Pressure Cell using water and lysis solution (containing detergents and chaotropic agents) and the Mickle beating of the cells in the presence of lysis solution and water. In this study, the standard lysis solution containing urea and CHAPS was used as it was shown to be compatible with different ProteinChips® for SELDI analysis (Cordingley *et al.,* 2003). Commonly used detergents such as SDS was shown to be unsuitable as it causes ion suppression (Barzaghi *et al.,* 2004). The extracts from the French Pressure Cell showed more protein bands on the 1D-gel than the extracts from the Mickle disintegration (Figure 3.20). The total protein content and the profiles on the SELDI also showed different results (Figure 3.19). Although the French Pressure Cell yielded a higher protein concentration (~ 30 µg/µl) and showed a higher number of peaks with a higher intensity in water extracts, Mickle beating combined with freeze-thawing with 30 µl lysostaphin produced an adequate amount (~ 10-15 µg/µl) of protein necessary for SELDI analysis and was the method of choice.

SELDI-TOF-MS and ProteinChip® arrays used in this study allowed rapid, high-throughput analysis of both large and small biomolecules with good reproducibility. Although the resulting spectra of MRSA and MSSA showed characteristic peaks specific to each strain, due to the vast amount of complex, noisy data generated by SELDI (approximately 13,000 data points per sample in the 3-30 kDa mass range), robust computer algorithms are vital to screen for potential biomarkers and remains a major challenge in bioinformatics (Ball *et al.,* 2002). To overcome this problem, the SELDI-TOF-MS data was analysed using a model based on Artificial Neural Network.

Artificial Neural Networks (ANNs) are a form of Artificial Intelligence (AI) capable of modelling for complex systems and fast emerging as one of the most popular tools for complex data analysis. It is a mathematical/ computational model based on biological neural networks that organizes and process information (Lancashire *et al.,* 2005). ANNs gathers knowledge by detecting patterns and relationships in data and learn through experience, not from programming. An ANN is formed from hundreds of single units, artificial neuron or processing elements (PE), connected with coefficients (weights), which constitute the neural structure and are organised in layers which resembles the biological neural network and therefore have many advantages (Kustrin and Beresford 2000). Firstly, they are essentially non-linear so that they are able to process data containing complex

interactions that are difficult to interpret. In addition, they are capable of generalisation, so they can interpret information which is different to that of the training data, thus representing a "real-world" solution to a given problem. A third advantage, which greatly helps proteomic data, is that they are fault tolerant, i.e. they have the ability of handling noisy or fussy information, while also being able to tolerate data which are incomplete or contain missing values. Like all other approaches, ANNs also have their limitations. The one major limitation of ANNs is that they do not explain how they reach a conclusion and often referred to as "black boxes" (Lancashire *et al.,* 2005). Also training of ANN can be time consuming depending on the complexity of the data. Furthermore, the quality of the input data also plays a major role in the quality of the output by the model.

There are many different network structures and different learning algorithms are applied ANNs (Abdi 2003). The most commonly used type of ANNs is the multi-layer perception (MLP) and back propagation (BP) algorithm because of its flexibility, adaptability and wide application capabilities (Wei *et al.,* 1998; Desilva *et al.,* 1994). In MLP, ANNs are organised into several layers, an input layer, one or more hidden layers and an output layer (Figure 4.3), with each layer having a number of neurons/ processing elements (Lancashire *et al.,* 2009). The input layer interacts with the external environment to receive the input data and output neurons provide ANNs response to the input data while hidden neurons communicate only with other neurons (Kustrin and Beresford 2000).

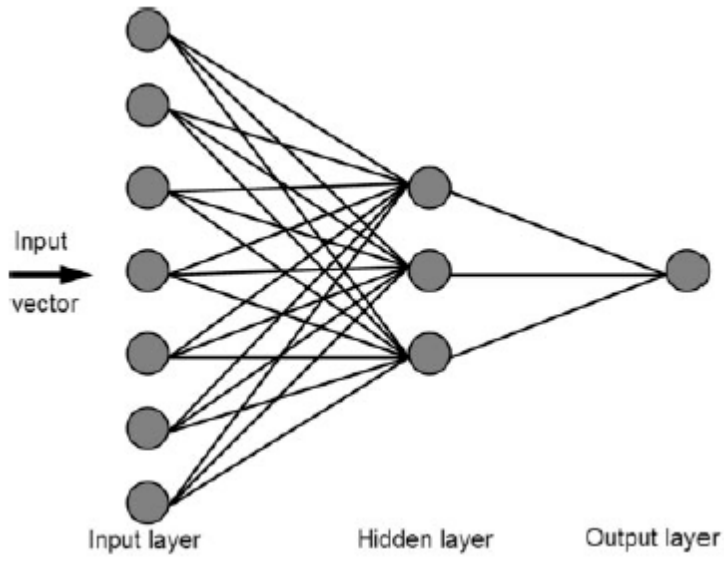


Figure 4.3: Architecture of a typical multi-layered perception ANN (Lancashire *et al.,* 2009)

During the development of the ANN model the data set was split into three sub sets; training data, test data and validation data. The ANN model was trained with the training data and continually optimised against the test data. The validation data was used to assess the ANN model on unseen data when the model was developed (Lancashire *et al.,* 2005). Once an ANN model is built for a particular application, it should be trained since ANNs gather its knowledge through a learning process. There are two major learning algorithms used in ANNs; supervised and unsupervised. Supervised learning involves providing the network with a set of cases that have values for the input as well as the known desired outputs (Lancashire *et al.,* 2009). In unsupervised learning, input information is only used for the training purpose, as the network itself governs how it groups the cases based on the input data. ANNs learn through special training algorithms such as BP algorithms (Lancashire *et al.,* 2005). It contains two main steps, a forward step which involves generating a solution to the problem and the back propagation of the error to modify the weights in order to reach a minimum error. However, if a network is left to train for too long, it will over train and will lose the ability to generalise (Kustrin and Beresford 2000). When the ANN produced the desired output after training, the weighted links between the units are saved. These weights are then used as analytical tools to predict results for a new set of input data.

For this study, in order to determine whether ANNs have the capability to identify specific biomarker ions, analysis was carried out using a stepwise approach in order to rank the ions based on the ability to predict as MRSA and MSSA. Hence unimportant and noisy values were removed and remaining single ions (3000-30,000 Da) were fed into the ANN model. From the first set of inputs, ion 3081 Da was chosen as the best predictive ion because of it's lower mean error and compared to the rest of the ions and the process was repeated until the most important sub set of ions were achieved which was between 3000-19,000 *m/z* values. After choosing the seven most predictive ions (3081, 5709, 5893, 7694, 9580, 15308 and 18896 Da), each ion was reanalysed for predictive performance using 50 ANN models.

The predictive performance of the model based on the seven most distinct ions in the population distribution curve revealed that all of the MRSA isolates (except one) were correctly assigned by these ions while only two MSSA isolates were misclassified as MRSA (Figure 3.27). This may have been due to the loss of its resistance to methicillin or they maybe in the process of acquiring antibiotic resistance for example by acquiring

SCC*mec* type I, II and IV (Enright *et al.,* 2002) (Figure 1.2.). Apart from these inconsistencies, the results indicate that these seven ions could be possible biomarkers for MRSA as indicated by the error bars. The area under the curve value obtained from receiver operating curve (0.9147) also indicated that this model have a high sensitivity and specificity (91%). However, the model could be validated using a blind data set for further confirmation.

There are numerous applications of the ANNs technique in the field of bioscience, some of which have been applied to a number of human pathogens. A study carried out by Yang *et al.* (2009) for *S. aureus* and non-*S. aureus* isolates using SELDI-MS coupled with ANNs analysis, reported a sensitivity of 98.4 % and specificity of 98.6 % based on six protein peaks detected by the ANN model. Another study by Schmid *et al.* (2005) used to identify and characterise *Neisseria gonorrhoeae* and closely related species and achieved a sensitivity of 95.7 % and specificity of 97.1 % using ANNs. A recent study by Xiao *et al.* (2009) demonstrated the successful identification of *Klebsiella pneumoniae* with an accuracy of 96.9 % using ANNs. ANNs have also been used successfully for analysis of MADI-MS data for the separation of strains of *Bacillus anthracis* vegetative cells, mixtures of cells and spores (Lasch *et al.,* 2009).

SDS-PAGE was also used in this study with the aim of detecting specific biomarkers between MRSA and MSSA. When selecting a protein extraction method suitable for SDS-PAGE analysis, a boiling method (section 2.7.1 b) used with SDS was unable to produce adequate protein content, however, the addition of number of detergents and lysostaphin improved the yield. The NuPAGE gels of *S. aureus* showed similar protein bands in the molecular weight range of 75 kDa, 50 kDa 30 kDa and 15 kDa for both MRSA and MSSA while one of the MSSA band patterns (Figure 3.28) showed several unique peptide/protein bands. In order to detect the presence of specific proteins/peptides in MRSA (MW2) and MSSA (476), some of these specific bands were analysed by digesting with trypsin followed by MS/MS analysis. The MS/MS analysis was carried out on LTQ-Orbitrap MS and the peptides were analysed using a Mascot search (Matrix science Ltd.). The corresponding proteins were viewed using Scaffold 3 proteomic software. With reference to isolates MW2 (MRSA) and 476 (MSSA), 26 and 19 unique proteins were identified respectively indicating the differences between MRSA and MSSA isolates (Appendix VI). However, further studies involving more strains are necessary to confirm the presence of these proteins. A similar study by Silva Santos *et al.* (2009) also demonstrated the presence

of species-specific whole-cell protein profiles for 139 isolates and concluded that SDS-PAGE was useful for identifying clinically prevalent staphylococci species.

In an attempt to detect intracellular biomarkers and the investigate intraspecies diversity of *C. difficile*, a similar approach was carried out using SELDI and SDS-PAGE. A total of 114 isolates were analysed using SELDI and 28 by SDS-PAGE. The same protein extraction method was carried out as for *S. aureus* excluding lysozyme (Section 2.6.1). The protein extracts were analysed using three different ProteinChips® to select a suitable array. Since the profiles obtained using the Strong Anionic Exchange ProteinChip® array (SAX/Q10) showed a larger number of peaks with higher intensity, the Q10 array was chosen for further studies. The Q10 array contains positively charged, quaternary ammonium groups that interact with negative charges on the surface of target proteins.

The SELDI profiles of these isolates showed the presence of characteristic peaks/ biomarkers (~ 9500 and 13000 Da) while some of the isolates showed additional peaks (~ 3000-4500 and 14500 Da) on the MS profiles (Figure 3.31). This clearly shows the intraspecies diversity within *C. difficile*. Specifically comparing the SELDI-MS profiles of the three strains from 1970s- 2006 (T, B1 and L), additional peaks could be detected for strain T (1970s) and L (2006) in the mass range of ~ 3000-4500 Da, indicating changes may have occurred over the years (Figure 3.31). Interestingly the strain L, from the recent outbreak indicated additional peaks, which may correspond to its higher virulence. Further studies on these additional peaks may provide insight into its pathogenesis and SELDI-MS could also be used as a preliminary screening tool to detect the variability of the isolates.

The SDS-PAGE carried for these strains also showed very similar protein band profiles. However, most of the proteins were conserved within the range of 6 kDa and 14-17 kDa and ~38- 49 kDa region, several of which may be of S-layer proteins (Figure 3.32). S-layer proteins of *C. difficile* are thought to be one of the main virulence factors and believed to play a significant role in host-pathogen interactions (McCoubrey and Poxton 2001; Wright *et al.,* 2005). McCoubrey and Poxton (2001) also suggested the detection of S-layer proteins as a useful marker for phenotyping of *C. difficile* in epidemiological studies.

The results of this study and the vast number of recent publications (Seng *et al.,* 2009; Eigner *et al.,* 2009; Pennanec *et al.,* 2010; La Scola and Raoult 2010) now demonstrating the application of MALDI-TOF-MS within research and clinical laboratories indicate that

the technique is best used at the species level as demonstrated here. In microbiology laboratories, while it is necessary to first rapidly identify a pathogen to the species level, for tracing the path of a strain, for example during disease outbreak, high resolution typing methods are required. Because of the limitation of MALDI-TOF-MS to reach this level of resolution at the present time, a newly developing approach, VNTR was undertaken using a diverse collection of *C. difficile* isolates.

VNTR analysis of *C. difficile*

Epidemiological investigations require rapid, reliable and highly discriminatory genotyping methods to track transmission and identify the emergence of new variants of pathogenic species. Current genotyping methods for detection of *C. difficile* outbreaks and nosocomial transmission include PCR ribotyping, AP-PCR, AFLP, PFGE, MLST and REA (section 1.2.5). These methods are labour intensive, subjective or lack sufficient discriminatory power to differentiate between closely related isolates. Due to the drawbacks of these typing methods and in order to efficiently track and detect the transmission of *C. difficile* locally and globally, new genotyping methods are being investigated. This study focused on a genomic approach for characterising isolates of *C. difficile* with the emphasis being placed on two predominant ribotypes (027 and 001) of *C. difficile.*

A total of 92 *C. difficile* isolates belonging to different ribotypes were analysed to detect novel loci to differentiate *C. diffcile* by using Variable Number Tandem Repeats (VNTR). A total of 47 loci were tested for tandem repeats, including 10 loci that were published (Marsh *et al.,* 2006 and van den Berg *et al.,* 2007) (Table 2.11). The remaining 37 loci were identified from the sequenced *C. difficile* 630 genome using the Tandem Repeat Finder programme. Of the 47 loci, 8 loci including a published locus (CD16, 17, 26, 29, 31, 44, 47 and CD9) failed to give an amplification product from any isolate in the set and were excluded from analysis. The failure to amplify products may have been due to base pair mismatch at the 3' end of the forward or the reverse primer. The other reason could be that these repeats maybe absent in the isolates selected for this study. However, due to lack of time and the cost of designing new primers, further analysis on these primers was not considered. Also, a recent study by Marsh *et al.* (2010) omitted the use of CD9 primer in their study, as this locus generated few alleles and contains 2 tandem-repeat loci that cannot be resolved by automated capillary electrophoresis.

For all 45 ribotype 027 isolates, 14 of the 37 newly identified loci produced identical copy numbers (Table 3.4) and were unsuitable to use as markers to differentiate isolates within ribotype 027. Two of the new loci, CD14 and CD23 were detected with copy numbers 'null' and '0' respectively for ribotype 027 isolates. The ribotype 001 isolates and the remaining ribotypes yielded copy numbers different from those exhibited by ribotype 027 isolates. Hence, these two loci could be used as potential markers to differentiate ribotype 027 isolates from the rest of the ribotypes. In order to detect the corresponding genes for these particular loci, the amplification product was searched using BLAST (http://blast.ncbi.nlm.nih.gov/). The BLAST results indicated that CD14 and CD23 corresponds to toxin genes *tcd*B, which is responsible for the increased toxin production of toxin B in virulent 027 strains and *tcd*C, which is a putative regulatory factor of toxin A and B which leads to increased toxin production in ribotype 027 respectively (Stabler *et al.,* 2009). This confirmed the copy number variance detected in loci CD14 where 'null' result correspond to the 18bp deletion in *tcd*B gene in ribotype 027 isolates compared to ribotype 001 where the copy number was 2.

For cluster analysis, an arbitrary value of ≥ 90 % similarity was selected to detect the homogeneity among isolates. A study by Fawley *et al.* (2008) concluded that differences of > 71 % (up to 2 copy numbers) could be considered as closely related strains. In this study, cluster analysis of the ribotype 027 using the new loci revealed 7 groups containing isolates from the same outbreak with 100 % similarity. Interestingly among these isolates, one or two copy number differences could be observed up to two or more loci indicating the emergence of subtypes within a clone (Fawley *et al.,* 2008). A similar study by Zaiss *et al.* (2009) also observed the differences of variable copy numbers at 4 or more loci in isolates even when the isolates were affiliated to identical sequence types or ribotypes. However, according to stability tests performed by Marsh and van den Berg (2006 and 2007); a difference of one or two repeat units between strains should not be interpreted as separate types or subtypes. Hence, the isolates within each group could be considered as closely related isolates. For the published loci (CD4, CD5 and CD7) copy number variances were observed between this study and the study by Marsh and van den Berg (2006 and 2007). These differences could be due to sizing variation (up to three base pairs) observed in different electrophoresis instruments (Pasqualotto *et al.,* 2007). Hence, this could have been one of the reasons for the copy number variances observed in these two loci. One of the outbreak isolates, G49 clustered separately from the rest of the outbreak isolates and exhibited higher dissimilarity (53 %) on cluster analysis. This isolate was later

confirmed as a non-027 isolate by florescent ribotyping at the HPA, London. This indicates the high resolution power of the newly designed VNTR loci for differentiating ribotype 027 isolates from other ribotypes.

For ribotype 001 isolates, 10 loci (CD11, 12, 14, 18, 19, 20, 21, 22, 23 and 25) showed different copy numbers from ribotype 027 isolates. Hence, these could be potential markers for differentiating ribotype 001 isolates from 027 isolates (Table 3.7). BLAST results indicated that loci CD11, 19, 20, and 21 corresponds to CTn (putative conjugative transposon DNA recombination protein), CD12 corresponds to *tcd*C (putative regulatory factor), CD18 corresponds to *cdt*B (ADP-ribosyltransferase-pseudo gene), CD22 and CD23 corresponds to *tcd*C (putative exported protein) and CD25 corresponds to *Dan*R (putative antibiotic resistance ABC) in the *C. difficile* 630 genome. Similar to ribotype 027 results, the copy numbers for loci CD4 and CD6 obtained here deferred to that reported by van den Berg. For loci CD5, this study showed copy numbers of 5 or 6 whereas van den Berg *et al.* (2007) obtained a copy number of five. An identical copy number was exhibited by loci CD7 in both studies. This single copy number variance may have been due to the size variation of the fragments during the electrophoresis as described earlier (Pasqualotto *et al.,* 2007). Cluster analysis (Figure 3.36 and 3.37) using the new loci showed four groups each containing closely related isolates. However, cluster analysis using all 39 loci did not yield these groupings of closely related isolates. Within the four groups, more isolates showed 100 % similarity to each other although there were few copy number variances among the isolates as before. Except for two ribotype 001 isolates, G18 and G28, all the isolates were obtained from the Microbial Pathogencity Research Laboratory (MPRL), Edinburgh. Interestingly, the G28 isolate clustered with the MPRL isolates in group D (Figure 3.37).

For the 22 different ribotype isolates (13 known and 9 unknown ribotypes), all 39 loci showed variable copy numbers (Table 3.6). However, for loci CD18, 22 and 23 stable copy numbers of one, seven and one were detected respectively, independent of the ribotype. For the same loci, ribotype 027 isolates showed copy numbers of two, one and zero while ribotype 001 isolates showed copy numbers one, seven and one respectively. Hence, these three loci could be used as potential markers to differentiate other ribotypes including ribotype 001 from ribotypes 027. From the first cluster analysis (Figure 3.38) within ≥ 90 % similarity range, one group containing five isolates were observed. From this, two isolates belonged to ribotype 012 showed 100 % similarity to each other. However, from

the second cluster analysis performed by using the new VNTR loci, two groups were detected within the same similarity range (Figure 3.39). This indicates that the new loci have a higher discriminatory power to differentiate isolates of different ribotypes. Interestingly, in one group (B) (Figure 3.39), along with ribotype 012 isolates, a ribotype 002 isolate showed 100 % similarity indicating that these isolates could be closely related. Zaiss *et al.* (2009) detected a similar observation where two isolates were assigned to two different ribotypes with a summed tandem repeat difference of four repeats between the isolates and showed 100 % similarity to each other.

In order to assess the discriminatory power of the new loci, cluster analysis was performed for all different ribotype isolates. Interestingly, two main clusters were observed and the majority (90 %) of ribotype 027 isolates clustered into a single group and the second cluster contained the remaining isolates (Figure 3.40). Individual cluster analysis revealed 10 specific loci that could be used as potential markers to differentiate each of the three ribotype groups; ribotype 027, 001 and other ribotype groups (Table 3.7). When cluster analysis was performed using these 10 specific loci (CD11, 12, 14, 18, 19, 20, 21, 22, 23 and 25), the majority of isolates belonging to each of the three groups clustered separately from each other (Figure 3.41). This suggests that these 10 loci could be used as potential markers to differentiate ribotype 027, 001 and other ribotype isolates. However, there were outlier isolates within each group and these could be isolates that are closely related to other isolates. Alternatively, this problem can be resolved by the sequence analysis of the VNTR loci as sequence data provides accurate information at the DNA sequence level (Zaiss *et al.,* 2009) and could be used to detect phylogenetic diversity of each isolate.

The aim of this part of the study was to detect novel loci to differentiate within and between the prevalent ribotypes 027, 001 and other ribotypes. These loci could be used as potential markers to detect outbreak strains from circulating strains. According to the results of this study, it can be suggested that 10 new VNTR loci could be used to differentiate between isolates of different ribotypes belonging to 027, 001 and other ribotypes. This could potentially be used as a highly reproducible, high-throughput and discriminatory genotyping technique for the virulent ribotype 027 isolates and ribotype 001 of *C. difficile*. A similar study by Killgore *et al.* (2008) compared the discriminatory power of seven DNA fingerprinting techniques for 42 *C. difficile* isolates collected from four different countries, including 22 ribotype 027 isolates. They reported that only four methods MLVA, REA, *slpA* sequence typing and PFGE were able to recognise the

subtypes, while REA and MLVA had sufficient power to distinguish strains within type 027. Similar studies (Fawley *et al.,* 2008, Marsh *et al.,* 2006; van den Berg *et al.,* 2007; Zaiss *et al.,* 2009) have highlighted the discriminatory power of the VNTR/MLVA technique for different ribotypes of *C. difficile* isolates and its utility in understanding the nosocomial transmission and epidemiology of *C. difficile* within hospitals and institutions. One of the key advantages of this typing technique, as highlighted in many publications, is that this technique is easy to perform using multiplex PCR reactions and gives a real-time microbial fingerprinting method (Marsh *et al.,* 2006; Lindstedt 2005; van Belkum *et al.,* 2008). A major drawback of this technique is that the design of the loci is based on the genome of a single strain. Hence, designing universal primers might not be possible. This may have also been one of the reasons for the failure of amplification of products of the seven new loci. In addition, one of the important factors is that this typing technique should be applied to outbreaks to ensure the stability and reproducibility of each locus. Thus, further studies using these 10 new markers for other ribotypes of *C. difficile* should also be performed on a larger set of isolates. This could prove to be useful in finding evolutionary trends in *C. difficile*. The data should also be compared with other established typing methods and the epidemiological data of isolates.

As mass spectral techniques develop, it is likely that in the future MALDI-TOF-MS will also be used as a typing tool to support the above method. One such study by Hain (2010) was recently reported for *Listeria monocytogenes*. However at the present time, a genome (VNTR) – proteome (MALDI-TOF-MS) algorithm to identify and type a species represents a new approach that can be developed for any major pathogen. DNA-based methods are now well established, by contrast mass spectral approaches are relatively new. Notwithstanding this, it is evident that intact MALDI-TOF-MS has evolved as a rapid, high-throughput technique, which could be used as an alternative method in routine clinical laboratories. Rapid bacterial identification based on a standardised library of reference spectral profiles enables the accurate identification of bacteria up to species or even to sub-species level. Thus, evidence over the years has supported this view. However, the critical issue of the limited use of a database still restricts the broad application of this exciting technology.

In the future, an online database, similar to Genbank will enable Users of any MALDI mass spectrometer to undertake comparative analysis using this database and so broaden the applications of this technology. For example, the Waters® MMU database used earlier

in this study (Rajakaruna *et al.,* 2009) was restricted to Users who had acquired only Waters® instruments and the MicrobeLynx™ software while BioTyper is tied to Users with Bruker instruments. SARAMIS™ has more flexibility as it is a software company but it is necessary to obtain a licence for its use and to link it to a mass spectrometer. The development of an online database, analogous to that used for 16S rRNA, in which Users can freely access a database globally, will enable the technology to move to the forefront of the microbial identification and lead to a dramatic impact in diagnostic microbiology of human infections.

References

Aebersold, R. and M. Mann. 2003. Mass spectrometry-based proteomics. Nature **422**:198-207.

Abdi, H. 2003. Neural Networks. Encyclopedia of Social Sciences Research Methods. 1-5.

Adam, B.L., Q. Yinsheng., J. W. Davis., M. D. Ward., M. A. Clements., L. H. Cazares., O. J. Semmes., P. F. Schellhammer., Y. Yasui., Z. Feng., G. L. Wright Jr. 2001. Serum Protein Fingerprinting Coupled with a Pattern-matching Algorithm Distinguishes Prostate Cancer from Benign Prostate Hyperplasia and Healthy Men. Cancer Reaserch. **62**: 3609–3614.

Agatonovic-Kustrin, S. and R. Beresford. 2000. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. J.Pharm.Biomed.Anal. **22**:717-727.

Anderson, N.G and N. L. Anderson. 1996. Twenty years of two-dimensional electrophoresis: past, present and future. Electrophoresis. **17**:443-53.

Arnold, R. J. and J. P. Reilly. 1999. Observation of *Escherichia coli* ribosomal proteins and their posttranslational modifications by mass spectrometry. Anal.Biochem. **269**:105-112.

Austen, B. M., E. R. Frears., and H. Davies. 2000. The use of ProteinChip arrays to monitor production of Alzheimer's beta amyloid in transfected cells. J. Peptide Sci. **6**:459–469.

Ball, G., S. Mian, F. Holding, R. O. Allibone, J. Lowe, S. Ali, G. Li, S. McCardle, I. O. Ellis, C. Creaser, and R. C. Rees. 2002. An integrated approach utilizing artificial neural networks and SELDI mass spectrometry for the classification of human tumours and rapid identification of potential biomarkers. Bioinformatics. **18**:395-404.

Bantscheff, M., M. Schirle., G. Sweetman., J. Rick., B. Kuster. 2007. Quantitative mass spectrometry in proteomics: a critical review. Anal Bioanal Chem. **389**:1017–1031.

Barbuddhe, S. B., T. Maier, G. Schwarz, M. Kostrzewa, H. Hof, E. Domann, T. Chakraborty, and T. Hain. 2008. Rapid identification and typing of listeria species by matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl.Environ.Microbiol. **74**:5402-5407.

Barzaghi, D., J. D. Isbister, K. P. Lauer, and T. L. Born. 2004. Use of surface-enhanced laser desorption/ionization -time of flight to explore bacterial proteomes. Proteomics. **4**:2624-2628.

Bassetti, M., E. Nicco, and M. Mikulska. 2009. Why is community-associated MRSA spreading across the world and how will it change clinical practice? Int.J.Antimicrob.Agents **34 Suppl 1**:S15-S19.

Beavis, R.C and B.T. Chait. 1989. Cinnamic acid derivatives as matrices for ultraviolet laser desorption mass spectrometry of proteins. *Rapid*. Commun. Mass. Spectrom. **3**:432-435.

Bergey's Manual of Systematic Bacteriology. 2nd ed., 2009. Volume **3**: The Firmicutes.

Berger-Bachi, B. and S. Rohrer. 2002. Factors influencing methicillin resistance in *staphylococci*. Arch.Microbiol. **178**:165-171.

Bernardo, K., N. Pakulat, M. Macht, O. Krut, H. Seifert, S. Fleer, F. Hunger, and M. Kronke. 2002. Identification and discrimination of *Staphylococcus aureus* strains using matrix-assisted laser desorption/ionization-time of flight mass spectrometry. Proteomics. **2**:747-753.

Bhalla, A., J. Nicole., B.S. Pultz., M. D. M. Gries., A. J. Ray., E. C. Eckstein., D. C. Aron., C. J. Donskey. 2004. Acquisition of Nosocomial Pathogens on Hands After Contact With Environmental Surfaces Near Hospitalized Patients. Infection. Contol. Hos. Epidemiol. **25**:164-166.

Bidet, P., V. Lalande, B. Salauze, B. Burghoffer, V. Avesani, M. Delmee, A. Rossier, F. Barbut, and J. C. Petit. 2000. Comparison of PCR-ribotyping, arbitrarily primed PCR, and pulsed-field gel electrophoresis for typing *Clostridium difficile*. J.Clin.Microbiol. **38**:2484-2487.

Bizzini, A., C. Durussel, J. Bille, G. Greub, and G. Prod'hom. 2010. Performance of matrix-assisted laser desorption ionization-time of flight mass spectrometry for identification of bacterial strains routinely isolated in a clinical microbiology laboratory. J.Clin.Microbiol. **48**:1549-1554.

Blossom, D. B. and L. C. McDonald. 2007. The challenges posed by reemerging *Clostridium difficile* infection. Clin.Infect.Dis. **45**:222-227.

Borriello, S. P. and M. H. Wilcox. 1998. *Clostridium difficile* infections of the gut: the unanswered questions. J.Antimicrob.Chemother. **41 Suppl C**:67-69.

Borriello, S. P. 1998. Pathogenesis of *Clostridium difficile* infection. J.Antimicrob.Chemother. **41 Suppl C**:13-19.

Borriello, S.P., H.A. Davies., S. Kamiya., P.J. Reed., S. Seddon. 1990. Virulance factors of *Clostridium difficile*. Rev.of Infec.Disea. **12**:185-191.

Bowman, R.A., G.L. O'Neil., V.R. Thomas. 1991. Non-radioactive restriction fragment length polymorphism (RFLP) typing of *Clostridium difficile*. FEMS. Microbiol. Lett. **79**:269-272.

Bouza, E., P. Munoz, and R. Alonso. 2005. Clinical manifestations, treatment and control of infections caused by *Clostridium difficile*. Clin.Microbiol.Infect. **11 Suppl 4**:57-64.

Brazier, J. S. 1998. The diagnosis of *Clostridium difficile*-associated disease. J.Antimicrob.Chemother. **41 Suppl C**:29-40.

Brazier, J. S. 2001. Typing of *Clostridium difficile*. Clin.Microbiol.Infect. **7**:428-431.

Brazier, J. S., S. P. Borriello. 2000. Microbiology, Epidemiology and Diagnosis of *Clostridium difficile* Infection. Curr. Top. Microbiol. Immunol. **250**:1-33.

Bradford, M.M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal.Chem. **72**:248-255.

Breathnach, A. S. 2005. Nosocomial infections. Medicine. **33**:22-26.

Burdon, D. W. 1982. *Clostridium difficile*: the epidemiology and prevention of hospital-acquired infection. Infection. **10(4):**203-4.

Budowle, B., S. E. Schutzer, M. S. Ascher, R. M. Atlas, J. P. Burans, R. Chakraborty, J. J. Dunn, C. M. Fraser, D. R. Franz, T. J. Leighton, S. A. Morse, R. S. Murch, J. Ravel, D. L. Rock, T. R. Slezak, S. P. Velsko, A. C. Walsh, and R. A. Walters. 2005. Toward a system of microbial forensics: from sample collection to interpretation of evidence. Appl.Environ.Microbiol. **71**:2209-2213.

Calderón, Y.A., L. M. Contreras., V. R. Lemoine., E. Palacios-Pru. 2008. Antibiotic resistance patterns and SDS-PAGE protein profi les of Burkholderia cepacia complex isolates from nosocomial and environmental sources in Venezuela. Med. Sci. Monit. **14**(2): 49-55.

Canas, B., D. Lopez-Ferrer, A. Ramos-Fernandez, E. Camafeita, and E. Calvo. 2006. Mass spectrometry technologies for proteomics. Brief.Funct.Genomic.Proteomic. **4**:295-320.

Carbonnelle, E., J. L. Beretti, S. Cottyn, G. Quesne, P. Berche, X. Nassif, and A. Ferroni. 2007. Rapid identification of *Staphylococci* isolated in clinical microbiology laboratories by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J.Clin.Microbiol. **45**:2156-2161.

Casey, A. L., P. A. Lambert, and T. S. Elliott. 2007. *Staphylococci*. Int.J.Antimicrob.Agents **29 Suppl 3**:S23-S32.

Cato, E. P., George, W. L. & Finegold, S. M. (1986). Genus *Clostridium Prazmowski* 1880, 23. In Bergey's Manual of Systematic Bacteriology, vol. 2:1141-200. Edited by P. H. A. Sneath, N. S. Mair, M. E. Sharpe & J. G. Holt. Baltimore: Williams & Wilkins.

Cato, E. P., D. E. Hash, L. V. Holdeman, and W. E. Moore. 1982. Electrophoretic study of *Clostridium* species. J.Clin.Microbiol. **15**:688-702.

Celis, J. E. and P. Gromov. 1999. 2D protein electrophoresis: can it be perfected? Curr.Opin.Biotechnol. **10**:16-21.

Chambers, H. F. 2001. The changing epidemiology of *Staphylococcus aureus*? Emerg.Infect.Dis. **7**:178-182.

Chen, Y. D., S. Zheng, J. K. Yu, and X. Hu. 2004. Artificial neural networks analysis of surface-enhanced laser desorption/ionization mass spectra of serum protein pattern distinguishes colorectal cancer from healthy population. Clin.Cancer Res. **10**:8380-8385.

Chesney, P.J. 1981. Clinical manifestation of toxic shock syndrome. J. Amer. Med. Assn. **246**:741-748.

Cho, W. C. 2007. Proteomics technologies and challenges. Genomics Proteomics.Bioinformatics. **5**:77-85.

Chong, B. E., D. B. Wall., D. M. Lubman., S. J. Flynn. 1997. Rapid profiling of *E. coli* proteins up to 500 kDa from whole cell lysates using Matrix-Assisted Laser Desorption/Ionisation mass spectrometry. Rapid. Commun. Mas. Spectrum. **11**:1900-1908.

Clauser, K. R., P. Baker., A. L. Burlingame. 1999. Role of accurate mass measurement in protein identification strategies employing MS or MS/MS database searching. Anal. Chem. 2871-2882.

Clarridge, J. E., III. 2004. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. Clin.Microbiol.Rev. **17**:840-62, table.

Claydon, M. A., S. N. Davey, V. Edwards-Jones, and D. B. Gordon. 1996. The rapid identification of intact microorganisms using mass spectrometry. Nat.Biotechnol. **14** :1584-1586.

Cloud, J. and C. P. Kelly. 2007. Update on *Clostridium difficile* associated disease. Curr.Opin.Gastroenterol. **23**:4-9.

Collins, M. D., P. A. Lawson, A. Willems, J. J. Cordoba, J. Fernandez-Garayzabal, P. Garcia, J. Cai, H. Hippe, and J. A. Farrow. 1994. The phylogeny of the genus Clostridium: proposal of five new genera and eleven new species combinations. Int.J.Syst.Bacteriol. **44**:812-826.

Cordingley, H. C., S. L. Roberts, P. Tooke, J. R. Armitage, P. W. Lane, W. Wu, and S. E. Wildsmith. 2003. Multifactorial screening design and analysis of SELDI-TOF ProteinChip array optimization experiments. Biotechniques **34**:364-373.

Corriere, M. D. and C. F. Decker. 2008. MRSA: an evolving pathogen. Dis.Mon. **54**:751-755.

Costas, M., B. D. Cookson, H. G. Talsania, and R. J. Owen. 1989. Numerical analysis of electrophoretic protein patterns of methicillin-resistant strains of *Staphylococcus aureus*. J.Clin.Microbiol. **27**:2574-2581.

Costas, M., B. Holmes, A. C. Wood, and S. L. On. 1989. Numerical analysis of electrophoretic protein patterns of Providencia rettgeri strains from human faeces, urine and other specimens. J.Appl.Bacteriol. **67**:441-452.

Costas, M. 1995. Classification, identification and typing of bacteria by the analysis of their one-dimensional polyacrylmide gel enectrophoretic proein patterns. Methods in Mol. Biol. **46**:353-402.

Communicable Disease Surveillance Centre. 2000. *Clostridium difficile* in England and Wales: 1999. Commun. Disease Reports CDR Weekly **10**: 135.

Dai, Y., L. Li, D. C. Roser, and S. R. Long. 1999. Detection and identification of low-mass peptides and proteins from solvent suspensions of *Escherichia coli* by high performance liquid chromatography fractionation and matrix-assisted laser desorption/ionization mass spectrometry. Rapid Commun.Mass Spectrom. **13**:73-78.

Dailey, D. C., A. Kaiser, and R. H. Schloemer. 1987. Factors influencing the phagocytosis of *Clostridium difficile* by human polymorphonuclear leukocytes. Infect.Immun. **55**:1541-1546.

Das, I., N. O'Connell, and P. Lambert. 2007. Epidemiology, clinical and laboratory characteristics of *Staphylococcus aureus* bacteraemia in a university hospital in UK. J.Hosp.Infect. **65**:117-123.

Dawson, L. F., E. Valiente, and B. W. Wren. 2009. *Clostridium difficile*--a continually evolving and problematic pathogen. Infect.Genet.Evol. **9**:1410-1417.

Delmas, J., J. P. Chacornac, F. Robin, P. Giammarinaro, R. Talon, and R. Bonnet. 2008. Evaluation of the Vitek 2 system with a variety of *Staphylococcus* species. J.Clin.Microbiol. **46**:311-313.

Delmee, M. 2001. Laboratory diagnosis of *Clostridium difficile* disease. Clin.Microbiol.Infect. **7**:411-416.

Deneve, C., C. Janoir, I. Poilane, C. Fantinato, and A. Collignon. 2009. New trends in *Clostridium difficile* virulence and pathogenesis. Int.J.Antimicrob.Agents **33 Suppl 1**:S24-S28.

Desilva, C. J. S., P. I. Choong., and Y. Attikiouzel. 1994. Artificial Neural Networks and Brest Cancer prognosis. Aust. Comput. J. **26**:78-81.

Deurenberg, R. H., C. Vink, S. Kalenic, A. W. Friedrich, C. A. Bruggeman, and E. E. Stobberingh. 2007. The molecular evolution of methicillin-resistant *Staphylococcus aureus*. Clin.Microbiol.Infect. **13**:222-235.

Deurenberg, R. H. and E. E. Stobberingh. 2008. The evolution of *Staphylococcus aureus*. Infect.Genet.Evol. **8**:747-763.

Dieckmann, R., R. Helmuth, M. Erhard, and B. Malorny. 2008. Rapid classification and identification of salmonellae at the species and subspecies levels by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl.Environ.Microbiol. **74**:7767-7778.

Dieckmann, R., E. Strauch, and T. Alter. 2009. Rapid identification and characterization of *Vibrio* species using whole-cell MALDI-TOF mass spectrometry. J.Appl.Microbiol.

Dinges, M. M., P. M. Orwin, and P. M. Schlievert. 2000. Exotoxins of *Staphylococcus aureus*. Clin.Microbiol.Rev. **13**:16-34, table.

Dohnal, V., K. Kuca, and D. Jun. 2005. What are artificial neural networks and what they can do? Biomed.Pap.Med.Fac.Univ Palacky.Olomouc.Czech.Repub. **149**:221-224.

Domon, B. and R. Aebersold. 2006. Mass spectrometry and protein analysis. Science **312**:212-217.

Doshi, R. K., G. Patel, R. Mackay, and F. Wallach. 2009. Healthcare-associated Infections: epidemiology, prevention, and therapy. Mt.Sinai J.Med. **76**:84-94.

Drudy, D., N. Harnedy, S. Fanning, M. Hannan, and L. Kyne. 2007. Emergence and control of fluoroquinolone-resistant, toxin A-negative, toxin B-positive *Clostridium difficile.* Infect.Control Hosp.Epidemiol. **28**:932-940.

Du, Z., R. Yang, Z. Guo, Y. Song, and J. Wang. 2002. Identification of *Staphylococcus aureus* and determination of its methicillin resistance by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Anal.Chem. **74**:5487-5491.

Duim, B., J. A. Wagenaar, J. R. Dijkstra, J. Goris, H. P. Endtz, and P. A. Vandamme. 2004. Identification of distinct *Campylobacter lari* genogroups by amplified fragment length polymorphism and protein electrophoretic profiles. Appl.Environ.Microbiol. **70**:18-24.

Dupont, C., V. Sivadon-Tardy, E. Bille, B. Dauphin, J. L. Beretti, A. S. Alvarez, N. Degand, A. Ferroni, M. Rottman, J. L. Herrmann, X. Nassif, E. Ronco, and E. Carbonnelle. 2009. Identification of clinical coagulase-negative *staphylococci*, isolated in microbiology laboratories, by matrix-assisted laser desorption/ionization-time of flight mass spectrometry and two automated systems. Clin.Microbiol.Infect.

Easterling, M. L., C. M. Colangelo., R. A. Scott, I .J. Amster. 1998. Monitoring protein expression in whole bacterial cells with MALDI-time-of-flight mass spectrometry. Anal. Chem. **70**:2704-2709.

Edman, P. 1949. A method for the determination of amino acid sequence in peptides. Arch.Biochem. **22**:475.

Edwards-Jones, V., M. A. Claydon, D. J. Evason, J. Walker, A. J. Fox, and D. B. Gordon. 2000. Rapid discrimination between methicillin-sensitive and methicillin-resistant *Staphylococcus aureus* by intact cell mass spectrometry. J.Med.Microbiol. **49**:295-300.

Eigner, U., M. Holfelder, K. Oberdorfer, U. Betz-Wild, D. Bertsch, and A. M. Fahr. 2009. Performance of a matrix-assisted laser desorption ionization-time-of-flight mass spectrometry system for the identification of bacterial isolates in the clinical routine laboratory. Clin.Lab **55**:289-296.

Elrick, M. M., J. L. Walgren, M. D. Mitchell, and D. C. Thompson. 2006. Proteomics: recent applications and new technologies. Basic Clin.Pharmacol.Toxicol. **98**:432-441.

Emonet, S., H. N. Shah., C. Abdessalam., J. Schrenzel. Application and usage of various mass spectrometry methods in clinical microbiology. Clinical Microbiology and Infections (submitted May 2010).

Encheva, V. 2005. Proteome analysis of bacterial pathogens (PhD). University of East London.

Enright, M. C., N. P. Day, C. E. Davies, S. J. Peacock, and B. G. Spratt. 2000. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. J.Clin.Microbiol. **38**:1008-1015.

Enright, M. C., D. A. Robinson, G. Randle, E. J. Feil, H. Grundmann, and B. G. Spratt. 2002. The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA). Proc.Natl.Acad.Sci.U.S.A **99**:7687-7692.

Erhard, M., U. C. Hipler, A. Burmester, A. A. Brakhage, and J. Wostemeyer. 2008. Identification of dermatophyte species causing onychomycosis and tinea pedis by MALDI-TOF mass spectrometry. Exp.Dermatol. **17**:356-361.

Fawley, W. N., J. Freeman, C. Smith, C. Harmanus, R. J. van den Berg, E. J. Kuijper, and M. H. Wilcox. 2008. Use of highly discriminatory fingerprinting to analyze clusters of *Clostridium difficile* infection cases due to epidemic ribotype 027 strains. J.Clin.Microbiol. **46**:954-960.

Feng, Y., C. J. Chen, L. H. Su, S. Hu, J. Yu, and C. H. Chiu. 2008. Evolution and pathogenesis of *Staphylococcus aureus*: lessons learned from genotyping and comparative genomics. FEMS Microbiol.Rev. **32**:23-37.

Fenselau, C. and P. A. Demirev . 2001. Characterization of intact microorganisms by MALDI mass spectrometry. Mass Spectrom.Rev. **20**:157-171.

Flynn, N. and S. H. Cohen. 2008. The continuing saga of MRSA. J.Infect.Dis. **197**:1217-1219.

Fox, A. 2006. Mass spectrometry for species or strain identification after culture or without culture: Past, present, and future. J.Clin.Microbiol. **44**:2677-2680.

Frenay, H. M., A. E. Bunschoten, L. M. Schouls, W. J. van Leeuwen, C. M. Vandenbroucke-Grauls, J. Verhoef, and F. R. Mooi. 1996. Molecular typing of methicillin-resistant *Staphylococcus aureus* on the basis of protein A gene polymorphism. Eur.J.Clin.Microbiol.Infect.Dis. **15**:60-64.

Fung, E. T., V. Thulasiraman, S. R. Weinberger, and E. A. Dalmasso. 2001. Protein biochips for differential profiling. Curr.Opin.Biotechnol. **12**:65-69.

Garfin, D. E. 1990. One-Dimensional Gel Electrophoresis. Methods. Enzyme. **182**:425-441.

GeneMapper® Software v4.0 with Multi-User Client-Server Capability. (www.appliedbiosystems.com).

George, W. L., V. L. Sutter., D. Citron., S. M. Finegold. 1979. Selective and differential medium for isolation of *Clostridium difficile.* Infect. Immun. **69**:6004-6011.

Ghosh, R., G. B. Nair, L. Tang, J. G. Morris, N. C. Sharma, M. Ballal, P. Garg, T. Ramamurthy, and O. C. Stine . 2008. Epidemiological study of *Vibrio cholerae* using variable number of tandem repeats. FEMS Microbiol.Lett. **288**:196-201.

Giger, O., C. C. Charilaou, and K. R. Cundy. 1984. Comparison of the API Staph-Ident and DMS Staph-Trac systems with conventional methods used for the identification of coagulase-negative *staphylococci*. J.Clin.Microbiol. **19**:68-72.

Gleeson, T. D. 2008. Prevention and control of methicillin-resistant *Staphylococcus aureus*. Dis.Mon. **54**:801-806.

Goodfellow, M. 1995. Inter-strain comparison of pathogenic microrganisms by pyrolysis mass spectrometry. Binary.-Comp.Microbiol. **7:**54-60.

Gordon, R. J. and F. D. Lowy. 2008. Pathogenesis of methicillin-resistant *Staphylococcus aureus* infection. Clin.Infect.Dis. **46 Suppl 5**:S350-S359.

Gould, I. M. 2008. The epidemiology of antibiotic resistance. Int.J.Antimicrob.Agents **32 Suppl 1**:S2-S9.

Graham, R. L., C. Graham, and G. McMullan. 2007. Microbial proteomics: a mass spectrometry primer for biologists. Microb.Cell Fact. **6**:26.

Graves, P. R. and T. A. Haystead. 2002. Molecular biologist's guide to proteomics. Microbiol.Mol.Biol.Rev. **66**:39-63.

Grosse-Herrenthey, A., T. Maier, F. Gessler, R. Schaumann, H. Bohnel, M. Kostrzewa, and M. Kruger. 2008. Challenging the problem of *clostridial* identification with matrix-assisted laser desorption and ionization-time-of-flight mass spectrometry (MALDI-TOF MS). Anaerobe. **14**:242-249.

Gurtler, V. and B. C. Mayall. 2001. Genomic approaches to typing, taxonomy and evolution of bacterial isolates. Int.J.Syst.Evol.Microbiol. **51**:3-16.

Gurtler, V. 1993. Typing of *Clostridium difficile* strains by PCR-amplification of variable length 16S-23S rDNA spacer regions. J. General. Micobiol. **139**:3089-3097.

Hain, T. 2010. In: Proceedings of the European Congress of Clinical Microbiology and Infectious Diseases. Abstract No: **S225**.

Hantula, J., A. Kurki., P. Vuoriranta., D. H. Bamford. 1991. Rapid classification of bacterial strains by SDS-polyacrylamide gel electrophoresis: population dynamics of the dominant dispersed phase bacteria of activated sludge. Appl. Microbiol. Biotechnol. **34**:551-555.

Harrington, P., N. E. Vieira., J. Espinoza., J. K. Nien., R. Romeroc., A. L. Yergey. 2005 Analysis of variance–principal component analysis: A soft tool for proteomic discovery. Analytica. Chimica. Acta. **544**:118–127.

Haag, A. M., S. N. Taylor, K. H. Johnston, and R. B. Cole. 1998. Rapid identification and speciation of Haemophilus bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. J.Mass Spectrom. **33**:750-756.

Haddad, N., M. Masselot, and B. Durand. 2004. Molecular differentiation of Mycobacterium bovis isolates. Review of main techniques and applications. Res.Vet.Sci. **76**:1-18.

Hampel, D. J., C. Sansome, M. Sha, S. Brodsky, W. E. Lawson, and M. S. Goligorsky. 2001. Toward proteomics in uroscopy: urinary protein profiles after radiocontrast medium administration. J.Am.Soc.Nephrol. **12**:1026-1035.

Hardy, K. J., D. W. Ussery, B. A. Oppenheim, and P. M. Hawkey. 2004. Distribution and characterization of *staphylococcal* interspersed repeat units (SIRUs) and potential use for strain differentiation. Microbiology **150**:4045-4052.

Hatheway, C. L. 1990. Toxigenic *clostridia*. Clin.Microbiol.Rev. **3**:66-98.

Hennequin, C., F. Porcheray, A. Waligora-Dupriet, A. Collignon, M. Barc, P. Bourlioux, and T. Karjalainen. 2001. GroEL (Hsp60) of *Clostridium difficile* is involved in cell adherence. Microbiology **147**:87-96.

Hiramatsu, K., H. Hanaki, T. Ino, K. Yabuta, T. Oguri, and F. C. Tenover. 1997. Methicillin-resistant *Staphylococcus aureus* clinical strain with reduced vancomycin susceptibility. J.Antimicrob.Chemother. **40**:135-136.

Hiramatsu, K., L. Cui, M. Kuroda, and T. Ito. 2001. The emergence and evolution of methicillin-resistant *Staphylococcus aureus*. Trends Microbiol. **9**:486-493.

Hirsch, J., K. C. Hansen, A. L. Burlingame, and M. A. Matthay. 2004. Proteomics: current techniques and potential applications to lung disease. Am.J.Physiol Lung Cell Mol.Physiol **287**:L1-23.

Hilario, M., A. Kalousis., J. Prados., P. A. Binz. 2004. Data mining for mass spectra-based cancer diagnosis and biomarker discovery. DDT: BIOSILICO. **2**:214-222.

Hillenkamp, F., M. Karas., R. C. Beavis., B.T. Chait, 1991. Matrix-Assisted Laser Desorption/Ionisation mass spectrometry of biopolymers. Anal. Chem. **63**:1193A- 1203A.

Ho, P. L., C. W. Tse, G. C. Mak, K. H. Chow, and T. K. Ng. 2004. Community-acquired methicillin-resistant *Staphylococcus aureus* arrives in Hong Kong. J.Antimicrob.Chemother. **54**:845-846.

Holden, M. T., E. J. Feil, J. A. Lindsay, S. J. Peacock, N. P. Day, M. C. Enright, T. J. Foster, C. E. Moore, L. Hurst, R. Atkin, A. Barron, N. Bason, S. D. Bentley, C. Chillingworth, T. Chillingworth, C. Churcher, L. Clark, C. Corton, A. Cronin, J. Doggett, L. Dowd, T. Feltwell, Z. Hance, B. Harris, H. Hauser, S. Holroyd, K. Jagels, K. D. James, N. Lennard, A. Line, R. Mayes, S. Moule, K. Mungall, D. Ormond, M. A. Quail, E. Rabbinowitsch, K. Rutherford, M. Sanders, S. Sharp, M. Simmonds, K. Stevens, S. Whitehead, B. G. Barrell, B. G. Spratt, and J. Parkhill. 2004. Complete genomes of two clinical *Staphylococcus aureus* strains: evidence for the rapid evolution of virulence and drug resistance. Proc.Natl.Acad.Sci.U.S.A **101**:9786-9791.

Holland, R. D., J. G. Wilkes, F. Rafii, J. B. Sutherland, C. C. Persons, K. J. Voorhees, and J. O. Lay, Jr. 1996. Rapid identification of intact whole bacteria based on spectral patterns using matrix-assisted laser desorption/ionization with time-of-flight mass spectrometry. Rapid Commun.Mass Spectrom. **10**:1227-1232.

Huang, S. S., J. J. D. Diekema., D. K. Warren., G. Zuccotti., P. L. Winokur., S. Tendolkar., L. Boyken., R. Datta., R. M. Jones., M. A. Ward., T. Aubrey., A. B. Onderdonk., C. Garcia., R. Platt. 2008. Strain-Relatedness of Methicillin-Resistant *Staphylococcus aureus* Isolates Recovered from Patients with Repeated Infection. Clin Infect. Dis. **46**(8): 1241–1247.

Hutchens, T.W., T.U. Yip. 1993. New desorption strategies for the mass spectrometric analysis of macromolecules. Rapid Commun. Mas Spectrum.**7**:576-580.

Ines, I. 2008. Genomic and Proteomic analysis of *Staphylococcus aureus* including MRSA (PhD).

Iorio, N. L., R. B. Ferreira, R. P. Schuenck, K. L. Malvar, A. P. Brilhante, A. P. Nunes, C. C. Bastos, and K. R. dos Santos. 2007. Simplified and reliable scheme for species-level identification of *Staphylococcus* clinical isolates. J.Clin.Microbiol. **45**:2564-2569.

Issaq, H. J., T. D. Veenstra, T. P. Conrads, and D. Felschow. 2002. The SELDI-TOF MS approach to proteomics: protein profiling and biomarker identification. Biochem.Biophys.Res.Commun. **292**:587-592.

Ito, T., K. Okuma, X. X. Ma, H. Yuzawa, and K. Hiramatsu. 2003. Insights on antibiotic resistance of *Staphylococcus aureus* from its whole genome: genomic island SCC. Drug Resist.Updat. **6**:41-52.

Jarman, K. H., D. S. Daly, C. E. Petersen, A. J. Saenz, N. B. Valentine, and K. L. Wahl. 1999. Extracting and visualizing matrix-assisted laser desorption/ionization time-of-flight mass spectral fingerprints. Rapid Commun.Mass Spectrom. **13**:1586-1594.

Jevons, M.P. 1961. Celbenin-resistance *Staphylococci.* Br Med J. **1**:78-502.

Jeffreys, A. J., V. Wilson, and S. L. Thein. 1985. Individual-specific 'fingerprints' of human DNA. Nature **316**:76-79.

Kallow, W., M. Erhard., N. Lima., I. M. Santos., R. Serra., A. Venancio., T. Friedl., J. Muller., G. S. de Hoog and G. J. M. Verkley. 2006. Microbial strain characterisation by MALDI-TOF-MS possibilities and limits. Annual General Meeting of the European Culture Collections Organisation. 146-159.

Kaplan, M.H., and M. J. Tenenbaum. 1982. *Staphylococcus aureus*: cellular biology and clinical applications. Amer. J. Med.**72**:248-258.

Karas, M and F. Hillencamp 1988. Laser Desorption Ionisation of Proteins with Molecular Masses Exceeding 10000 Daltons. Anal. Chem. **60**:2299-2301.

Karas, M., D. Bachmann and F. Hillenkamp. 1985. Influence of the wavelength in High-Irradiance ultraviolet Laser Desorption Mass Spectrometry of organic molecules. Anal. Chem. **57**:2935-2939.

Kanduma, E., T. D. McHugh, and S. H. Gillespie. 2003. Molecular methods for *Mycobacterium tuberculosis* strain typing: a users guide. J.Appl.Microbiol. **94**:781-791.

Katayama, Y., T. Ito, and K. Hiramatsu. 2000. A new class of genetic element, *staphylococcus* cassette chromosome mec, encodes methicillin resistance in *Staphylococcus aureus*. Antimicrob.Agents Chemother. **44**:1549-1555.

Keys, C. J., D. J. Dare, H. Sutton, G. Wells, M. Lunt, T. McKenna, M. McDowall, and H. N. Shah. 2004. Compilation of a MALDI-TOF mass spectral database for the rapid

screening and characterisation of bacteria implicated in human infectious diseases. Infect.Genet.Evol. **4**:221-242.

Killgore, G., A. Thompson, S. Johnson, J. Brazier, E. Kuijper, J. Pepin, E. H. Frost, P. Savelkoul, B. Nicholson, R. J. van den Berg, H. Kato, S. P. Sambol, W. Zukowski, C. Woods, B. Limbago, D. N. Gerding, and L. C. McDonald. 2008. Comparison of seven techniques for typing international epidemic strains of *Clostridium difficile*: restriction endonuclease analysis, pulsed-field gel electrophoresis, PCR-ribotyping, multilocus sequence typing, multilocus variable-number tandem-repeat analysis, amplified fragment length polymorphism, and surface layer protein A gene sequence typing. J.Clin.Microbiol. **46**:431-437.

Kito, K. and T. Ito. 2008. Mass spectrometry-based approaches toward absolute quantitative proteomics. Curr.Genomics **9**:263-274.

Klevytska, A. M., L. B. Price, J. M. Schupp, P. L. Worsham, J. Wong, and P. Keim. 2001. Identification and characterization of variable-number tandem repeats in the *Yersinia pestis* genome. J.Clin.Microbiol. **39**:3179-3185.

Kloos, W. E. and J. F. Wolfshohl. 1982. Identification of *Staphylococcus* species with the API STAPH-IDENT system. J.Clin.Microbiol. **16**:509-516.

Kloos, W. E. 1990. Systematics and the natural history of *staphylococci* 1. Soc.Appl.Bacteriol.Symp.Ser. **19**:25S-37S.

Kloos, W. E. and C. G. George . 1991. Identification of *Staphylococcus* species and subspecies with the MicroScan Pos ID and Rapid Pos ID panel systems. J.Clin.Microbiol. **29**:738-744.

Kloos, W. E., D. N. Ballard, C. G. George, J. A. Webster, R. J. Hubner, W. Ludwig, K. H. Schleifer, F. Fiedler, and K. Schubert. 1998. Delimiting the genus *Staphylococcus* through description of *Macrococcus caseolyticus* gen. nov., comb. nov. and *Macrococcus equipercicus* sp. nov., and *Macrococcus bovicus* sp. no. and *Macrococcus carouselicus* sp. nov. Int.J.Syst.Bacteriol. **48 Pt 3**:859-877.

Kolker, E., R. Higdon, and J. M. Hogan. 2006. Protein identification and expression analysis using mass spectrometry. Trends Microbiol. **14**:229-235.

Koreen, L., S. V. Ramaswamy, E. A. Graviss, S. Naidich, J. M. Musser, and B. N. Kreiswirth. 2004. spa typing method for discriminating among *Staphylococcus aureus* isolates: implications for use of a single marker to detect genetic micro- and macrovariation. J.Clin.Microbiol. **42**:792-799.

Krishnamurthy, T. and P. L. Ross. 1996. Rapid identification of bacteria by direct matrix-assisted laser desorption/ionization mass spectrometric analysis of whole cells. Rapid Commun.Mass Spectrom. **10**:1992-1996.

Kuijper, E. J., J. H. Oudbier, W. N. Stuifbergen, A. Jansz, and H. C. Zanen. 1987. Application of whole-cell DNA restriction endonuclease profiles to the epidemiology of *Clostridium difficile*-induced diarrhea. J.Clin.Microbiol. **25**:751-753.

Kuijper, E. J., B. Coignard, and P. Tull. 2006. Emergence of *Clostridium difficile*-associated disease in North America and Europe. Clin.Microbiol.Infect. **12 Suppl 6**:2-18.

Kustrin, S and R. Beresford. 2000. Basic concepts of artificial neural network (ANN) modelling and its application in pharmaceutical research. J. Pharm. Biomed. Anal. **22**: 717–727.

Kyne, L., M. B. Hamel, R. Polavaram, and C. P. Kelly. 2002. Health care costs and mortality associated with nosocomial diarrhea due to *Clostridium difficile*. Clin.Infect.Dis. **34**:346-353.

Ladhani, S., C. L. Joannou, D. P. Lochrie, R. W. Evans, and S. M. Poston. 1999. Clinical, microbial, and biochemical aspects of the exfoliative toxins causing *staphylococcal* scalded-skin syndrome. Clin.Microbiol.Rev. **12**:224-242.

Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227**:680-685.

Lancashire, L., O. Schmid, H. Shah, and G. Ball. 2005. Classification of bacterial species from proteomic data using combinatorial approaches incorporating artificial neural networks, cluster analysis and principal components analysis. Bioinformatics. **21**:2191-2199.

Lancashire, L. J., C. Lemetre, and G. R. Ball. 2009. An introduction to artificial neural networks in bioinformatics--application to complex microarray and mass spectrometry datasets in cancer studies. Brief.Bioinform. **10**:315-329.

La Scola, B., D. Raoult. 2010. Direct Identification of Bacteria in Positive Blood Culture Bottles by Matrix-Assisted Laser Desorption Ionisation Time-of-Flight Mass Spectrometry. PLoS ONE. **4**:1-6.

Lauber, W. M., J. A. Carroll., D. R. Dufield., J. R. Kiesel., M. R. Radabaugh., J. P. Malone. 2001. Mass spectrometry compatibility of two-dimensional gel protein stains. Electrophoresis. **22**: 906–918.

Lasch, P., W. Beyer, H. Nattermann, M. Stammler, E. Siegbrecht, R. Grunow, and D. Naumann. 2009. Identification of Bacillus anthracis by using matrix-assisted laser desorption ionization-time of flight mass spectrometry and artificial neural networks. Appl.Environ.Microbiol. **75**:7229-7242.

Lay, J. O., Jr. 2001. MALDI-TOF mass spectrometry of bacteria. Mass Spectrom.Rev. **20**:172-194.

Layer, F., B. Ghebremedhin, K. A. Moder, W. Konig, and B. Konig. 2006. Comparative study using various methods for identification of *Staphylococcus* species in clinical specimens. J.Clin.Microbiol. **44**:2824-2830.

Liang, X., K. Zheng., M. G. Qian., D.M. Lubman. 1996. Determination of bacterial protein profiles by Matrix-Assisted Laser Desorption/Ionisation mass spectrometry with high performance liquid chromatography. Rapid. Commun. Mas. Spectrum. **10:**1219-1226.

Lin, D., D. L. Tabb, and J. R. Yates, III. 2003. Large-scale protein identification using mass spectrometry. Biochim.Biophys.Acta **1646**:1-10.

Lina, G., Y. Piemont, F. Godail-Gamot, M. Bes, M. O. Peter, V. Gauduchon, F. Vandenesch, and J. Etienne. 1999. Involvement of Panton-Valentine leukocidin

producing *Staphylococcus aureus* in primary skin infections and pneumonia. Clin.Infect.Dis. **29**:1128-1132.

Lindsay, J. A., A. Ruzin, H. F. Ross, N. Kurepina, and R. P. Novick. 1998. The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus*. Mol.Microbiol. **29**:527-543.

Lindstedt, B. A., E. Heir, E. Gjernes, and G. Kapperud. 2003. DNA fingerprinting of Salmonella enterica subsp. enterica serovar typhimurium with emphasis on phage type DT104 based on variable number of tandem repeat loci. J.Clin.Microbiol. **41**:1469-1479.

Lindstedt, B. A. 2005. Multiple-locus variable number tandem repeats analysis for genetic fingerprinting of pathogenic bacteria. Electrophoresis **26**:2567-2582.

Lista, F., G. Faggioni, S. Valjevac, A. Ciammaruconi, J. Vaissaire, C. le Doujet, O. Gorge, R. De Santis, A. Carattoli, A. Ciervo, A. Fasanella, F. Orsini, R. D'Amelio, C. Pourcel, A. Cassone, and G. Vergnaud. 2006. Genotyping of *Bacillus anthracis* strains based on automated capillary 25-loci multiple locus variable-number tandem repeats analysis. BMC.Microbiol. **6**:33.

Livermore, D. M. 2000. Antibiotic resistance in *staphylococci*. Int.J.Antimicrob.Agents **16 Suppl 1**:S3-10.

Loo, V. G., L. Poirier, M. A. Miller, M. Oughton, M. D. Libman, S. Michaud, A. M. Bourgault, T. Nguyen, C. Frenette, M. Kelly, A. Vibien, P. Brassard, S. Fenn, K. Dewar, T. J. Hudson, R. Horn, P. Rene, Y. Monczak, and A. Dascal. 2005. A predominantly clonal multi-institutional outbreak of *Clostridium difficile*-associated diarrhea with high morbidity and mortality. N.Engl.J.Med. **353**:2442-2449.

Lowy, F. D. 1998. *Staphylococcus aureus* infections. N.Engl.J.Med. **339**:520-532.

Lyerly, D. M., H. C. Krivan, and T. D. Wilkins. 1988. *Clostridium difficile*: its disease and toxins. Clin.Microbiol.Rev. **1**:1-18.

Lynn, E. C., M. C. Chung, W. C. Tsai, and C. C. Han. 1999. Identification of Enterobacteriaceae bacteria by direct matrix-assisted laser desorptiom/ionization mass spectrometric analysis of whole cells. Rapid Commun.Mass Spectrom. **13**:2022-2027.

Maier, T., B. Wegemann., S. Schubert. In: Proceedings of the European Congress of Clinical Microbiology and Infectious Diseases. Abstract No: **S535**.

Marouga, R., S. David., E. Hawkins. 2005. The development of the DIGE system: 2D fluorescence difference gel analysis technology. Anal. Bioanal. Chem. **382**:669-678.

Magee, J. T., J. S. Brazier, I. K. Hosein, C. D. Ribeiro, D. W. Hill, A. Griffiths, C. Da Costa, A. J. Sinclair, and B. I. Duerden. 1993. An investigation of a nosocomial outbreak of *Clostridium difficile* by pyrolysis mass spectrometry. J.Med.Microbiol. **39**:345-351.

Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc.Natl.Acad.Sci.U.S.A **95**:3140-3145.

Marklein, G., M. Josten, U. Klanke, E. Muller, R. Horre, T. Maier, T. Wenzel, M. Kostrzewa, G. Bierbaum, A. Hoerauf, and H. G. Sahl. 2009. Matrix-assisted laser desorption ionization-time of flight mass spectrometry for fast and reliable identification of clinical yeast isolates. J.Clin.Microbiol. **47**:2912-2917.

Marsh, J. W., M. M. O'Leary, K. A. Shutt, A. W. Pasculle, S. Johnson, D. N. Gerding, C. A. Muto, and L. H. Harrison. 2006. Multilocus variable-number tandem-repeat analysis for investigation of *Clostridium difficile* transmission in Hospitals. J.Clin.Microbiol. **44**:2558-2566.

Marsh, J. W., M. M. O'Leary, K. A. Shutt, S. P. Sambol, S. Johnson, D. N. Gerding, and L. H. Harrison. 2010. Multilocus variable-number tandem-repeat analysis and multilocus sequence typing reveal genetic relationships among *Clostridium difficile* isolates genotyped by restriction endonuclease analysis. J.Clin.Microbiol. **48**:412-418.

Mazzeo, M. F., A. Sorrentino, M. Gaita, G. Cacace, M. Di Stasio, A. Facchiano, G. Comi, A. Malorni, and R. A. Siciliano. 2006. Matrix-assisted laser desorption ionization-time of flight mass spectrometry for the discrimination of food-borne microorganisms. Appl.Environ.Microbiol. **72**:1180-1189.

McMillin, D.E., L.L. Muldrow. 1992. Typing of toxic strains of *Clostridium difficile* using DNA fingerprints generated with arbitrary polymerace chain reaction primers. FEMS. Microbiol. Lett. **92**:5-10.

McCoubrey, J. and I. R. Poxton. 2001. Variation in the surface layer proteins of *Clostridium difficile*. FEMS Immunol.Med.Microbiol. **31**:131-135.

McDonald, L. C., G. E. Killgore, A. Thompson, R. C. Owens, Jr., S. V. Kazakova, S. P. Sambol, S. Johnson, and D. N. Gerding. 2005. An epidemic, toxin gene-variant strain of *Clostridium difficile*. N.Engl.J.Med. **353**:2433-2441.

McDonald, L. C., M. Owings, and D. B. Jernigan. 2006. *Clostridium difficile* infection in patients discharged from US short-stay hospitals, 1996-2003. Emerg.Infect.Dis. **12**:409-415.

McDougal, L. K., C. D. Steward, G. E. Killgore, J. M. Chaitram, S. K. McAllister, and F. C. Tenover. 2003. Pulsed-field gel electrophoresis typing of oxacillin-resistant *Staphylococcus aureus* isolates from the United States: establishing a national database. J.Clin.Microbiol. **41**:5113-5120.

McFarland, L. V. 2006. Meta-analysis of probiotics for the prevention of antibiotic associated diarrhea and the treatment of *Clostridium difficile* disease. Am.J.Gastroenterol. **101**:812-822.

Melles, D. C., R. F. Gorkink, H. A. Boelens, S. V. Snijders, J. K. Peeters, M. J. Moorhouse, P. J. van der Spek, W. B. van Leeuwen, G. Simons, H. A. Verbrugh, and A. van Belkum. 2004. Natural population dynamics and expansion of pathogenic clones of *Staphylococcus aureus*. J.Clin.Invest **114**:1732-1740.

Melles, D. C., W. B. van Leeuwen, S. V. Snijders, D. Horst-Kreft, J. K. Peeters, H. A. Verbrugh, and A. van Belkum. 2007. Comparison of multilocus sequence typing (MLST), pulsed-field gel electrophoresis (PFGE), and amplified fragment length

polymorphism (AFLP) for genetic typing of *Staphylococcus aureus*. J.Microbiol.Methods **69**:371-375.

Mendes, P. 2002. Emerging bioinformatics for the metabolome. Brief.Bioinform. **3**:134-145.

Merchant, M. and S. R. Weinberger. 2000. Recent advancements in surface-enhanced laser desorption/ionization-time of flight-mass spectrometry. Electrophoresis **21**:1164-1177.

Mirza, S. P. and M. Olivier. 2008. Methods and approaches for the comprehensive characterization and quantification of cellular proteomes using mass spectrometry. Physiol Genomics **33**:3-11.

Moos, M., Jr., N. Y. Nguyen, and T. Y. Liu. 1988. Reproducible high yield sequencing of proteins electrophoretically separated and transferred to an inert support. J.Biol.Chem. **263**:6005-6008.

Mulligan, M. E., L. R. Peterson, R. Y. Kwok, C. R. Clabots, and D. N. Gerding. 1988. Immunoblots and plasmid fingerprints compared with serotyping and polyacrylamide gel electrophoresis for typing *Clostridium difficile*. J.Clin.Microbiol. **26**:41-46.

Murchan, S., M. E. Kaufmann, A. Deplano, R. de Ryck, M. Struelens, C. E. Zinn, V. Fussing, S. Salmenlinna, J. Vuopio-Varkila, N. El Solh, C. Cuny, W. Witte, P. T. Tassios, N. Legakis, W. van Leeuwen, A. van Belkum, A. Vindel, I. Laconcha, J. Garaizar, S. Haeggman, B. Olsson-Liljequist, U. Ransjo, G. Coombes, and B. Cookson. 2003. 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. J.Clin.Microbiol. **41**:1574-1585.

Murchan, S., H. M. Aucken, G. L. O'Neill, M. Ganner, and B. D. Cookson. 2004. Emergence, spread, and characterization of phage variants of epidemic methicillin-resistant *Staphylococcus aureus* 16 in England and Wales. J.Clin.Microbiol. **42**:5154-5160.

Muto, C. A., M. Pokrywka, K. Shutt, A. B. Mendelsohn, K. Nouri, K. Posey, T. Roberts, K. Croyle, S. Krystofiak, S. Patel-Brown, A. W. Pasculle, D. L. Paterson, M. Saul, and L. H. Harrison. 2005. A large outbreak of *Clostridium difficile*-associated disease with an unexpected proportion of deaths and colectomies at a teaching hospital following increased fluoroquinolone use. Infect.Control Hosp.Epidemiol. **26**:273-280.

Nakamura, S., T. Serikawa., M. Mikawa., S. Nakashio., K. Yamakawa., S. Nishida. 1981. Agglutination, Toxigenicity and Sorbitol fermentation of *Clostridium difficile.* Microbiol. Immunol. **25**:863-870.

Nelson, L, C. S. Cockram., G. Lui., R. Lam., E. Lam., R. Lai., M. Ip. 2006. Community case of methicillin-resistant *Staphylococcus aureus* infection. Emerg Infect Dis. Jan; **12**(1):172-4.

Nilsson, C. L. 1999. Fingerprinting of Helicobacter pylori strains by matrix-assisted laser desorption/ionization mass spectrometric analysis. Rapid Commun.Mass Spectrom. **13**:1067-1071.

Noller, A. C., M. C. McEllistrem, K. A. Shutt, and L. H. Harrison. 2006. Locus-specific mutational events in a multilocus variable-number tandem repeat analysis of *Escherichia coli* O157:H7. J.Clin.Microbiol. **44**:374-377.

NuPAGE® Technical Guide. Version E **:** October 2003.

O'Connor, D., P. Hynes, M. Cormican, E. Collins, G. Corbett-Feeney, and M. Cassidy. 2001. Evaluation of methods for detection of toxins in specimens of feces submitted for diagnosis of *Clostridium difficile*-associated diarrhea. J.Clin.Microbiol. **39**:2846-2849.

O'Farrell, P. H. 1975. High resolution two-dimensional electrophoresis of proteins. J.Biol.Chem. **250**:4007-4021.

Overman, T. L. and J. K. Overley. 1990. Reproducibility of API Staph-Ident system identifications of coagulase-negative *staphylococci* isolated from blood. J.Clin.Microbiol. **28**:2585-2586.

Patti, J. M., B. L. Allen., M. J. McGavin., M. Hook. 1994. MSCRAMM-mediated adherence of microorganisms to host tissues. Annu. Rev. Microbiol. **48**:585-617.

Pasqualotto, A. C., D. W. Denning, and M. J. Anderson. 2007. A cautionary tale: Lack of consistency in allele sizes between two laboratories for a published multilocus microsatellite typing system. J.Clin.Microbiol. **45**:522-528.

Peacock, S. J., G. D. de Silva, A. Justice, A. Cowland, C. E. Moore, C. G. Winearls, and N. P. Day. 2002. Comparison of multilocus sequence typing and pulsed-field gel electrophoresis as tools for typing *Staphylococcus aureus* isolates in a microepidemiological setting. J.Clin.Microbiol. **40**:3764-3770.

Pennanec, X., A. Dufour, D. Haras, and K. Rehel. 2010. A quick and easy method to identify bacteria by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. Rapid Commun.Mass Spectrom. **24**:384-392.

Poxton, I. R., J. McCoubrey, and G. Blair. 2001. The pathogenicity of *Clostridium difficile*. Clin.Microbiol.Infect. **7**:421-427.

Pineda, F.J., M.D. Antoine, P.A. Demirev, A.B. Feidman, J. Jackman, M. Longenecker, J.S. Lin. 2003. Microorganism identification by Martix-Assisted-Lase/Desorption Ionisation mass spectrometry and model-derived ribosomal protein biomarkers. Anal. Chem. **75**:3817-3822.

Prod'hom, G. A. Bizzini., C. Durussel., J. Bille., G. Greub. In: Proceedings of the European Congress of Clinical Microbiology and Infectious Diseases. Abstract No: **S536**.

Rajakaruna, L., G. Hallas, L. Molenaar, D. Dare, H. Sutton, V. Encheva, R. Culak, I. Innes, G. Ball, A. M. Sefton, M. Eydmann, A. M. Kearns, and H. N. Shah. 2009. High throughput identification of clinical isolates of *Staphylococcus aureus* using MALDI-TOF-MS of intact cells. Infect.Genet.Evol. **9**:507-513.

Ramazanzadeh, R. and R. McNerney. 2007. Variable Number Of Tandem Repeats (VNTR) and its application in bacterial epidemiology. Pak.J.Biol.Sci. **10**:2612-2621.

Reyes, R. C., M. A. John, D. L. Ayotte, A. Covacich, S. Milburn, and Z. Hussain. 2007. Performance of TechLab C. DIFF QUIK CHEK and TechLab C. DIFFICILE TOX A/B II for the detection of Clostridium difficile in stool samples. Diagn.Microbiol.Infect.Dis. **59**:33-37.

Rhoads, S., L. Marinelli, C. A. Imperatrice, and I. Nachamkin. 1995. Comparison of MicroScan WalkAway system and Vitek system for identification of gram-negative bacteria. J.Clin.Microbiol. **33**:3044-3046.

Robbins, J. B., R. Schneerson, G. Horwith, R. Naso, and A. Fattom. 2004. *Staphylococcus aureus* types 5 and 8 capsular polysaccharide-protein conjugate vaccines. Am.Heart J. **147**:593-598.

Roe, M. R. and T. J. Griffin. 2006. Gel-free mass spectrometry-based high throughput proteomics: tools for studying biological response of proteins and proteomes. Proteomics. **6**:4678-4687.

Ruelle, V., B. El Moualij, W. Zorzi, P. Ledent, and E. D. Pauw. 2004. Rapid identification of environmental bacterial strains by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun.Mass Spectrom. **18**:2013-2019.

Rupnik, M., M. H. Wilcox, and D. N. Gerding. 2009. *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. Nat.Rev.Microbiol. **7**:526-536.

Russmann, H., K. Panthel, R. C. Bader, C. Schmitt, and R. Schaumann. 2007. Evaluation of three rapid assays for detection of *Clostridium difficile* toxin A and toxin B in stool specimens. Eur.J.Clin.Microbiol.Infect.Dis. **26**:115-119.

Ryzhov, V. and C. Fenselau. 2001. Characterization of the protein subset desorbed by MALDI from whole bacterial cells. Anal.Chem. **73**:746-750.

Sabat, A., J. Krzyszton-Russjan, W. Strzalka, R. Filipek, K. Kosowska, W. Hryniewicz, J. Travis, and J. Potempa. 2003. New method for typing *Staphylococcus aureus* strains: multiple-locus variable-number tandem repeat analysis of polymorphism and genetic relationships of clinical isolates. J.Clin.Microbiol. **41**:1801-1804.

Said-Salim, B., B. Mathema, and B. N. Kreiswirth. 2003. Community-acquired methicillin-resistant *Staphylococcus aureus*: an emerging pathogen. Infect.Control Hosp.Epidemiol. **24**:451-455.

Saunders, N. A., A. Underwood, A. M. Kearns, and G. Hallas. 2004. A virulence-associated gene microarray: a tool for investigation of the evolution and pathogenic potential of *Staphylococcus aureus*. Microbiology **150**:3763-3771.

Schindler, C. A. and V. T. Schuhardt. 1964. Lysostaphin: A new bacteriolytic agent for the *staphylococcus*. Proc.Natl.Acad.Sci.U.S.A **51**:414-421.

Schmid, O., G. Ball, L. Lancashire, R. Culak, and H. Shah. 2005. New approaches to identification of bacterial pathogens by surface enhanced laser desorption/ionization time of flight mass spectrometry in concert with artificial neural networks, with special reference to *Neisseria gonorrhoeae*. J.Med.Microbiol. **54**:1205-1211.

Sebaihia, M., B. W. Wren, P. Mullany, N. F. Fairweather, N. Minton, R. Stabler, N. R. Thomson, A. P. Roberts, A. M. Cerdeno-Tarraga, H. Wang, M. T. Holden, A. Wright, C. Churcher, M. A. Quail, S. Baker, N. Bason, K. Brooks, T. Chillingworth, A. Cronin, P. Davis, L. Dowd, A. Fraser, T. Feltwell, Z. Hance, S. Holroyd, K. Jagels, S. Moule, K. Mungall, C. Price, E. Rabbinowitsch, S. Sharp, M. Simmonds, K. Stevens, L. Unwin, S. Whithead, B. Dupuy, G. Dougan, B. Barrell, and J. Parkhill. 2006. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. Nat.Genet. **38**:779-786.

Seddon, S. V., I. Hemingway, and S. P. Borriello. 1990. Hydrolytic enzyme production by *Clostridium difficile* and its relationship to toxin production and virulence in the hamster model. J.Med.Microbiol. **31**:169-174.

Seibert, V., A. Wiesner, T. Buschmann, and J. Meuer. 2004. Surface-enhanced laser desorption ionization time-of-flight mass spectrometry (SELDI TOF-MS) and ProteinChip technology in proteomics research. Pathol.Res.Pract. **200**:83-94.

Seibold, E., R. Bogumil, S. Vorderwulbecke, S. Al Dahouk, A. Buckendahl, H. Tomaso, and W. Splettstoesser. 2007. Optimized application of surface-enhanced laser desorption/ionization time-of-flight MS to differentiate *Francisella tularensis* at the level of subspecies and individual strains. FEMS Immunol.Med.Microbiol. **49**:364-373.

Sell, T. L., D. R. Schaberg, and F. R. Fekety. 1983. Bacteriophage and bacteriocin typing scheme for *Clostridium difficile*. J.Clin.Microbiol. **17**:1148-1152.

Seng, P., M. Drancourt, F. Gouriet, B. La Scola, P. E. Fournier, J. M. Rolain, and D. Raoult. 2009. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin.Infect.Dis. **49**:543-551.

Shah, H. N., C. J. Keys, O. Schmid, and S. E. Gharbia. 2002. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and proteomics: a new era in anaerobic microbiology. Clin.Infect.Dis. **35**:S58-S64.

Shah, H. N., L. Molenaar, L. Rajakaruna, J. E. Russell, D. Dare, H. Sutton, and G. Hallas. 2008. Assessment of the stability of cell-surface components of microorganisms by MALDI-TOF-MS following preservation on lenticule discs. FEMS Microbiol.Lett. **285**:284-290.

Shah, H. N and S. E. Gharbia. Mass spectrometry for microbial proteomics. 2010: Wiley.

Shah, H. N., C. J. Keys., S. E. Gharbia., K. Ralphson., F. Trundle., I. Brookhouse., M. A. Claydon. 2000. The application of Matrix-Assisted Laser Desorption/Ionisation Time of Flight Mass Spectrometry to profile the surface of Intact Bacterial Cells. Microb. Ecol. **12**:241-246.

Shah, H. N., V. Encheva., O. Schmid., R. A. Culak., P. Nasir., I. Ines., M. A. Chattaway., C. J. Keys., R. C. Jacinto.m L. Molenaar., R. S. Ayanza., G. Hallas., J. V. Hookey., D. Rajemdram. 2005. SELDI-TOF-MS: A potentially powerful tool for rapid characterisation of microorganisms. Encyclopedia of Rapid Microbiological Methods. Vol. III. DHI publishing.

Shopsin, B. and B. N. Kreiswirth. 2001. Molecular epidemiology of methicillin-resistant *Staphylococcus aureus*. Emerg.Infect.Dis. **7**:323-326.

Silva Santos, O. C., M. S. Laport, L. M. Teixeira, M. Giambiagi-Demarval, N. L. Iorio, and K. R. dos Santos. 2009. Reliable identification of clinically prevalent species and subspecies of *staphylococci* by sodium dodecyl sulfate polyacrylamide gel electrophoresis analysis. Diagn.Microbiol.Infect.Dis. **64**:1-5.

Siuzdak, G. 1994. The emergence of mass spectrometry in biochemical research. Proc.Natl.Acad.Sci.U.S.A **91**:11290-11297.

Snell, H., M. Ramos, S. Longo, M. John, and Z. Hussain. 2004. Performance of the TechLab C. DIFF CHEK-60 enzyme immunoassay (EIA) in combination with the C. difficile Tox A/B II EIA kit, the Triage *C. difficile* panel immunoassay, and a cytotoxin assay for diagnosis of Clostridium difficile-associated diarrhea. J.Clin.Microbiol. **42**:4863-4865.

Sousa, M. A., H. de Lencastre. 2004. Bridges from hospitals to the laboratory: genetic portraits of methicillin-resistant *Staphylococcus aureus* clones. FEMS Immunology and Medical Microbiology. **40**:101-111.

Stoica, G. E., A. Kuo., A. Aigner. 2001. Identification of anaplastic lymphoma kinase as a receptor for the growth factor pleiotrophin. J. Biol. Chem. **276**:16772–16779.

Stabler, R. A., M. He, L. Dawson, M. Martin, E. Valiente, C. Corton, T. D. Lawley, M. Sebaihia, M. A. Quail, G. Rose, D. N. Gerding, M. Gibert, M. R. Popoff, J. Parkhill, G. Dougan, and B. W. Wren. 2009. Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. Genome Biol. **10**:R102.

Starr, J. 2005. *Clostridium difficile* associated diarrhoea: diagnosis and treatment. BMJ **331**:498-501.

Stoddart, B. and M. H. Wilcox . 2002. *Clostridium difficile*. Curr.Opin.Infect.Dis. **15**:513-518.

Strand, M., T. A. Prolla, R. M. Liskay, and T. D. Petes. 1993. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature **365**:274-276.

Stubbs, S. L., J. S. Brazier, G. L. O'Neill, and B. I. Duerden. 1999. PCR targeted to the 16S-23S rRNA gene intergenic spacer region of *Clostridium difficile* and construction of a library consisting of 116 different PCR ribotypes. J.Clin.Microbiol. **37**:461-463.

Supply, P., E. Mazars, S. Lesjean, V. Vincent, B. Gicquel, and C. Locht. 2000. Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome. Mol.Microbiol. **36**:762-771.

Swenson, J. M., P. P. Williams., G. Killgore., C. M. O'Hara., F. C. Tenover. 2001. Performance of Eight Methods, Including Two New Rapid Methods, for Detection of Oxacillin Resistance in a Challenge Set of *Staphylococcus aureu*s Organisms. J. Clin. Microbiol. **39**(10):3785-8.

Tanaka, K., H. Waki., Y. Ido., S., Akita., Y. Yoshida., T. Yoshida. 1988. Protein and Polymer analysis up to m/z 100 000 by laser ionisation Time-of- Flight Mass Spectrometry. Rapid Commun. Mas. Spectrum. 151-153.

Tanaka, K., Y. Ido., S. Akita., Y. Yoshida., T. Yoshida. 1987. Detection of highmass molecules by laser desorption time-of-flight mass spectrometry. The mass spec. society. 185-188.

Taylor, A. J., M. Costas, and R. J. Owen. 1987. Numerical analysis of electrophoretic protein patterns of *Bacteroides ureolyticus* clinical isolates. J.Clin.Microbiol. **25**:660-666.

Teasley, D. G., D. N. Gerding, M. M. Olson, L. R. Peterson, R. L. Gebhard, M. J. Schwartz, and J. T. Lee, Jr. 1983. Prospective randomised trial of metronidazole versus vancomycin for *Clostridium-difficile*-associated diarrhoea and colitis. Lancet **2**:1043-1046.

Tenover, F. C., R. Arbeit, G. Archer, J. Biddle, S. Byrne, R. Goering, G. Hancock, G. A. Hebert, B. Hill, R. Hollis, and . 1994. Comparison of traditional and molecular methods of typing isolates of *Staphylococcus aureus*. J.Clin.Microbiol. **32**:407-415.

Tenover, F. C., R. D. Arbeit, R. V. Goering, P. A. Mickelsen, B. E. Murray, D. H. Persing, and B. Swaminathan . 1995. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. J.Clin.Microbiol. **33**:2233-2239.

Tenover, F. C. and R. V. Goering. 2009. Methicillin-resistant *Staphylococcus aureus* strain USA300: origin and epidemiology. J.Antimicrob.Chemother. **64**:441-446.

Thulasiraman, V., S. L. Mc-Cutchen-Maloney., V. L. Motin., E. Garcia. 2001. Detection and Identification of Virulance Fsctors in *Yersinia pestis* using SELDI ProteinChip® system. Biotec. **30**:428-432.

Torres, J. F. 1991. Purification and characterisation of toxin B from a strain of *Clostridium difficile* that does not produce toxin A. J.Med.Microbiol. **35**:40-44.

Tyers, M. and M. Mann. 2003. From genomics to proteomics. Nature **422**:193-197.

van Belkum, A. 2007. Tracing isolatesof bacterial species by multilocus variable number of tandem repeat analysis (MLVA). FEMS. Immunol. Med. Microbiol. **49**: 22–27.

van Belkum, A. 2008. Newer Methods for Bacterial Strain Typing. Clinical Microbiol. Newsletter. **30**:63-69.

van Belkum, A., S. Scherer, W. van Leeuwen, D. Willemse, L. van Alphen, and H. Verbrugh. 1997. Variable number of tandem repeats in clinical strains of *Haemophilus influenzae*. Infect.Immun. **65**:5017-5027.

van Belkum, A., S. Scherer, L. van Alphen, and H. Verbrugh. 1998. Short-sequence DNA repeats in prokaryotic genomes. Microbiol.Mol.Biol.Rev. **62**:275-293.

van den Berg, R. J., I. Schaap, K. E. Templeton, C. H. Klaassen, and E. J. Kuijper. 2007. Typing and subtyping of *Clostridium difficile* isolates by using multiple-locus variable-number tandem-repeat analysis. J.Clin.Microbiol. **45**:1024-1028.

Vamdamme, P., H. P. Endtz., J. S. Vliegenthart., H. W. Weverink., N. P. Van den Berg., H. A. Verbrugh., A. Van Belkum. 1997. Genotypic diversity of *Campylobacter lari* isolated from mussels and oysters in The Netherlands. Int. J. Food Microb. **34**: 79-88.

Vandamme, P., U. Torck, E. Falsen, B. Pot, H. Goossens, and K. Kersters. 1998. Whole-cell protein electrophoretic analysis of viridans *streptococci*: evidence for heterogeneity among *Streptococcus mitis* biovars. Int.J.Syst.Bacteriol. **48 Pt 1**:117-125.

Vandenesch, F., T. Naimi, M. C. Enright, G. Lina, G. R. Nimmo, H. Heffernan, N. Liassine, M. Bes, T. Greenland, M. E. Reverdy, and J. Etienne. 2003. Community-acquired methicillin-resistant *Staphylococcus aureus* carrying Panton-Valentine leukocidin genes: worldwide emergence. Emerg.Infect.Dis. **9**:978-984.

Vanlaere, E., K. Sergeant, P. Dawyndt, W. Kallow, M. Erhard, H. Sutton, D. Dare, B. Devreese, B. Samyn, and P. Vandamme. 2008. Matrix-assisted laser desorption ionisation-time-of of-flight mass spectrometry of intact cells allows rapid identification of *Burkholderia cepacia* complex. J.Microbiol.Methods **75**:279-286.

Vlahou, A., P. F. Schellhammer, S. Mendrinos, K. Patel, F. I. Kondylis, L. Gong, S. Nasim, and J. G. Wright, Jr. 2001. Development of a novel proteomic approach for the detection of transitional cell carcinoma of the bladder in urine. Am.J.Pathol. **158**:1491-1502.

Walker, J., A. J. Fox, V. Edwards-Jones, and D. B. Gordon. 2002. Intact cell mass spectrometry (ICMS) used to type methicillin-resistant *Staphylococcus aureus*: media effects and inter-laboratory reproducibility. J.Microbiol.Methods **48**:117-126.

Wang, Z., L. Russon, L. Li, D. C. Roser, and S. R. Long. 1998. Investigation of spectral reproducibility in direct analysis of bacteria proteins by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun.Mass Spectrom. **12**:456-464.

Warny, M., J. Pepin, A. Fang, G. Killgore, A. Thompson, J. Brazier, E. Frost, and L. C. McDonald. 2005. Toxin production by an emerging strain of *Clostridium difficile* associated with outbreaks of severe disease in North America and Europe. Lancet **366**:1079-1084.

Wei, J, T., Z. Zang., S. D. Barnhil., K. R. Madyastha., H. Zhang., J. E. Oesterlin. 1998. Understanding Artificial Neural Networks and exploring their potential applications for the practising urologist. Urology. **52**:161-72.

Welham, K. J., M. A. Domin, D. E. Scannell, E. Cohen, and D. S. Ashton. 1998. The characterization of micro-organisms by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun.Mass Spectrom. **12**:176-180.

Weller, T. M. 2000. Methicillin-resistant *Staphylococcus aureus* typing methods: which should be the international standard? J.Hosp.Infect. **44**:160-172.

Wenzel, R. P. and T. M. Perl. 1995. The significance of nasal carriage of *Staphylococcus aureus* and the incidence of postoperative wound infection. J.Hosp.Infect. **31**:13-24.

Wielders, C. L., M. R. Vriens, S. Brisse, L. A. Graaf-Miltenburg, A. Troelstra, A. Fleer, F. J. Schmitz, J. Verhoef, and A. C. Fluit. 2001. In-vivo transfer of mecA DNA to *Staphylococcus aureus* [corrected]. Lancet **357**:1674-1675.

Wilcox, M. H., J. G. Cunniffe, C. Trundle, and C. Redpath. 1996. Financial burden of hospital-acquired *Clostridium difficile* infection. J.Hosp.Infect. **34**:23-30.

Wilcox, M.H. 2003. *Clostridium difficile* infection and pseudomembranous colitis. Best Practice & Research Clinical Gastroenterology. **17**:475-493.

Wilkins, M. R., A. A. Scanchez., R. D. Gooley., I. Appel., D. F. Humphery-Smith., Hochstrasser and K. L. Williams. 1995. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. Biotechnol. Genet. Eng. Rev. **13:**19-50.

Wilkins, T. D. and D. M. Lyerly. 2003. *Clostridium difficile* testing: after 20 years, still challenging. J.Clin.Microbiol. **41**:531-534.

Williams, T. L., D. Andrzejewski, J. O. Lay, and S. M. Musser. 2003. Experimental factors affecting the quality and reproducibility of MALDI TOF mass spectra obtained from whole bacteria cells. J.Am.Soc.Mass Spectrom. **14**:342-351.

Winslow, C. E., J. Broadhurst, R. E. Buchanan, C. Krumwiede, L. A. Rogers, and G. H. Smith. 1920. The Families and Genera of the Bacteria: Final Report of the Committee of the Society of American Bacteriologists on Characterization and Classification of Bacterial Types. J.Bacteriol. **5**:191-229.

Wright, A., R. Wait., S. Begum., B. Crossett., J. Nagy., K. Brown., N. Fairweather. 2005. Proteomic analysis of cell surface proteins from *Clostridium difficile*. Proteomics. **5**:2443–2452.

Woo, P. C., S. K. Lau, J. L. Teng, H. Tse, and K. Y. Yuen. 2008. Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. Clin.Microbiol.Infect. **14**:908-934.

Wu, J. A., C. Kusuma, J. J. Mond, and J. F. Kokai-Kun. 2003. Lysostaphin disrupts Staphylococcus aureus and *Staphylococcus epidermidis* biofilms on artificial surfaces. Antimicrob.Agents Chemother. **47**:3407-3414.

Wu, S. W., H. de Lencastre, and A. Tomasz. 2001. Recruitment of the mecA gene homologue of *Staphylococcus sciuri* into a resistance determinant and expression of the resistant phenotype in *Staphylococcus aureus*. J.Bacteriol. **183**:2417-2424.

Wust, J., N. M. Sullivan, U. Hardegger, and T. D. Wilkins. 1982. Investigation of an outbreak of antibiotic-associated colitis by various typing methods. J.Clin.Microbiol. **16**:1096-1101.

Wulfkuhle, J. D., K. C. McLean., C. P. Pawaletz., D. C. Sgroi., B. J. Trock., P. S. Steeg and E. F. Petricoin. 2001. New approaches to proteomic analysis of breast cancer. Proteomics. **1:**1205–1215.

Xiao, D., Y. Yang, H. Liu, H. Yu, Y. Yan, W. Huang, W. Jiang, W. Liao, Q. Hu, and B. Huang. 2009. Development of a method based on surface enhanced laser desorption and ionization time of flight mass spectrometry for rapid identification of *Klebsiella pneumoniae*. J.Microbiol. **47**:646-650.

Xu, G., C. Q. Xiang, Y. Lu, X. N. Kang, P. Liao, Q. Ding, and Y. F. Zhang. 2009. Application of SELDI-TOF-MS to identify serum biomarkers for renal cell carcinoma. Cancer Lett. **282**:205-213.

Yang, S. Y., X. Y. Xiao, W. G. Zhang, L. J. Zhang, W. Zhang, B. Zhou, G. Chen, and D. C. He. 2005. Application of serum SELDI proteomic patterns in diagnosis of lung cancer. BMC.Cancer **5**:83.

Yang, Y. C., H. Yu, D. W. Xiao, H. Liu, Q. Hu, B. Huang, W. J. Liao, and W. F. Huang. 2009. Rapid identification of *Staphylococcus aureus* by surface enhanced laser desorption and ionization time of flight mass spectrometry. J.Microbiol.Methods **77**:202-206.

Zaiss, N. H., M. Rupnik, E. J. Kuijper, C. Harmanus, D. Michielsen, K. Janssens, and U. Nubel. 2009. Typing *Clostridium difficile* strains based on tandem repeat sequences. BMC.Microbiol. **9**:6.

Zhou, M. and T. Veenstra. 2008. Mass spectrometry: m/z 1983-2008. Biotechniques **44**:667-8, 670.

Appendices

Appendix I

Comparison of the two MALDI-MS studies of *S. aureus* done at HPA and MMU using database versions 2005 and 2006 respectively.

HPA No.	Hospital ID.	First Run ID by HPA	MMU ID	Second run for incorrect samples ID by MMU at HPA	16S ID
HPA 26	MRSA	*S.aureus*	*S.aureus*		
HPA 30	MRSA	*S.aureus*	*S.aureus*		
HPA 39	MRSA	*S.aureus*	*S.aureus*		
HPA 40	MRSA	*S.aureus*	*S.aureus*		
HPA 41	MRSA	*S.aureus*	*S.aureus*		
HPA 75	MRSA	*S.aureus*	*S.aureus*		
HPA 76	MRSA	*S.aureus*	*S.aureus*		
HPA 77	MRSA	*S.aureus*	*S.aureus*		
HPA 78	MRSA	*S.aureus*	*S.aureus*		
HPA 79	MRSA	*S.aureus*	*S.aureus*		
HPA 80	MRSA	*S. pyaogenus*	*S.aureus*	*S.aureus*	*S.aureus*
HPA 81	MRSA	*S.aureus*	*S.aureus*		
HPA 82	MRSA	*S.aureus*	*S.aureus*		
HPA 83	MRSA	*S.aureus*	*S.aureus*		
HPA 84	MRSA	*S.aureus*	*S.aureus*		
HPA 85	MRSA	*S.aureus*	*S.aureus*		
HPA 86	MRSA	*S.aureus*	*S.aureus*		
HPA 133	MRSA	*S.aureus*	*S.aureus*		
HPA 134	MRSA	*S.aureus*	*S.aureus*		
HPA 139	MRSA	*S.aureus*	*S.aureus*		
HPA 140	MRSA	*S.aureus*	*S.aureus*		
HPA 144	MRSA	*S.aureus*	*S.aureus*		
HPA 160	MRSA	*S.aureus*	*S.aureus*		
HPA 222	MRSA	*S.aureus*	*S.aureus*		
HPA 229	MRSA	*S.aureus*	*S.aureus*		
HPA 230	MRSA	*S.aureus*	*S.aureus*		
HPA 233	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 239	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 242	MRSA	*S.aureus*	*S.aureus*		
HPA 247	MRSA	*S.aureus*	*S.aureus*		
HPA 248	*S.aureus*	*S.aureus*	*Proteus mirabilis* 1-8	*S.aureus*	
HPA 249	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 250	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 256	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 257	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 258	MRSA	*S.aureus*	*S.aureus*		
HPA 259	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 260	*S.aureus*	*S.aureus*	1-*S. warneri* / *S.aureus* 2,3,4,6,7	*S.aureus*	
HPA 261	MRSA	S.aureus	*S.aureus*		
HPA 262	MRSA	S.aureus	*S.aureus*		
HPA 273	*S.aureus*	S.aureus	*S.aureus*		
HPA 279	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 280	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 281	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 284	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 285	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 287	*S.aureus*	*S.aureus*	*S.aureus*		

HPA No.	Hospital ID.	First Run ID by HPA	MMU ID	Second run for incorrect samples ID by MMU at HPA	16S ID
HPA 293	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 294	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 299	*S.aureus*	*S.aureus*	Inconclusive	*S.aureus*	
HPA 318	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 319	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 322	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 323	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 324	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 334	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 339	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 344	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 345	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 346	MRSA	*S.aureus*	*S.aureus*		
HPA 347	MRSA	*S.aureus*	*S.aureus*		
HPA 348	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 353	MRSA	*S.aureus*	*S.aureus*		
HPA 356	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 358	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 389	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 401	MRSA	*S.aureus*	*S.aureus*		
HPA 404	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 405	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 408	MRSA	*S.aureus*	*S.aureus*		
HPA 409	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 410	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 412	MRSA	*S.aureus*	*S.aureus*		
HPA 441	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 442	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 443	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 444	MRSA	*S.aureus*	*S.aureus*		
HPA 489	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 496	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 497	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 499	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 500	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 501	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 523	MRSA	*S.aureus*	*S.aureus*		
HPA 524	MRSA	*S.aureus*	*S.aureus*		
HPA 545	MRSA	*S.aureus*	*S.aureus*		
HPA 546	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 547	*S.aureus*	*S.haemolyticus*	*S.aureus*		*S.aureus*
HPA 549	*S.aureus*	*S.epidermidis*	Missing from the storage and couldn't send		*S.aureus*
HPA 550	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 556	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 563	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 569	*S.aureus*	*S.aureus*	*Pseudomonas aeruginosa*	*S.aureus*	
HPA 571	*S.aureus*	*S.aureus*	*S.aureus*		
HPA 573	MRSA	*S.aureus*	*S.aureus*		

Appendix II

MALDI-MS results of *S. aureus* isolates (n= 39) received from Staphylococcal Reference Laboratory.

All isolates were identified as *S. aureus* by the software apart from the isolates labelled in red, were confirmed from the second run.

Sample ID	Source	MALDI ID
H072660527	Clinical isolate	*S. aureus*
H072580475	Clinical isolate	*S. aureus*
H071920422	Clinical isolate	*S. aureus*
H071200363	Clinical isolate	*S. aureus*
H072940574	Clinical isolate	*S. aureus*
H073020460	Clinical isolate	*S. aureus*
H072660333	Clinical isolate	*S. aureus*
H072800374	Clinical isolate	*S. aureus*
H072340414	Clinical isolate	*S. aureus*
H072640555	Clinical isolate	*S. aureus*
H072500404	Clinical isolate	*S. aureus*
H072240800	Clinical isolate	*S. aureus*
H072900348	Clinical isolate	*S. aureus*
H072580470	Clinical isolate	*S. aureus*
H064560442	Clinical isolate	*S. aureus*
H072860414	Clinical isolate	*S. aureus*
H041940150	Clinical isolate	*S. aureus*
H072680434	Clinical isolate	*S. aureus*
H072640554	Clinical isolate	*S. aureus*
H072280533	Clinical isolate	*S. aureus*
H072820462	Clinical isolate	*S. aureus*
H072860467	Clinical isolate	*S. aureus*
H072580469	Clinical isolate	*S. aureus*
H072820481	Clinical isolate	*S. aureus*
H072740472	Clinical isolate	*S. aureus*
H072240582	Clinical isolate	*S. aureus*
H072820446	Clinical isolate	*S. aureus*
H072720523	Clinical isolate	*S. aureus*
H072680446	Clinical isolate	*S. aureus*
H072660541	Clinical isolate	*S. aureus*
H072860465	Clinical isolate	*S. aureus*
H072720526	Clinical isolate	*S. aureus*
H072820480	Clinical isolate	*S. aureus*
RH070000211	Clinical isolate	*S. aureus*
RH070000253	Clinical isolate	*S. aureus*
H072920482	Clinical isolate	*S. aureus*
H072660333	Clinical isolate	*S. aureus*
H072820464	Clinical isolate	*S. aureus*
H071380629	Clinical isolate	*S. aureus*

Appendix III

SARAMIS™ identification of 42 (n= 42) isolates subcultured on CBA for 24 h analysed in duplicates. MPRL 4857 was identified as *S. warneri* due to the contamination of the sample. Isolate MPRL 307 was not identified from the 24 h culture as *C. difficile*. But, was confirmed from the 48 h culture.

Strain ID	Culture media	Incubation time	SARAMIS ID	Confidence level (%)
MPRL 002	CBA	24 h	*C. difficile*	96.4 91.0
MPRL 059	CBA	24 h	*C. difficile*	99.9 91.0
MPRL 1037	CBA	24 h	*C. difficile*	96.4 81.0
MPRL 13366	CBA	24 h	*C. difficile*	90.0 89.0
MPRL 199	CBA	24 h	*C. difficile*	89.0 99.9
MPRL 223	CBA	24 h	*C. difficile*	99.9 99.9
MPRL 2282	CBA	24 h	*C. difficile*	95.0 95.0
MPRL 2783	CBA	24 h	*C. difficile*	99.9 99.9
MPRL 279	CBA	24 h	*C. difficile*	99.9 99.9
MPRL 282	CBA	24 h	*C. difficile*	96.4 96.4
MPRL 296	CBA	24 h	*C. difficile*	99.9 99.9
MPRL 307	CBA	24 h	No ID	No spectrum No spectrum
MPRL 308	CBA	24 h	*C. difficile*	81.0 96.4
MPRL 371	CBA	24 h	*C. difficile*	95.0 89.0
MPRL 372	CBA	24 h	*C. difficile*	91.0 96.4
MPRL 379	CBA	24 h	*C. difficile*	95.0 95.0
MPRL 382	CBA	24 h	*C. difficile*	96.4 No spectrum
MPRL 396	CBA	24 h	*C. difficile*	89.0 91.0
MPRL 402	CBA	24 h	*C. difficile*	81.0 99.9
MPRL 405	CBA	24 h	*C. difficile*	95.0 89.0
MPRL 406	CBA	24 h	*C. difficile*	89.0 91.0
MPRL 407	CBA	24 h	*C. difficile*	90.0 89.0

MPRL 418	CBA	24 h	*C. difficile*	99.9 95.0
MPRL 422	CBA	24 h	*C. difficile*	No spectrum 91.0
MPRL 4846	CBA	24 h	*C. difficile*	96.4 96.4
MPRL 4847	CBA	24 h	*C. difficile*	99.9 95.0
MPRL 4848	CBA	24 h	*C. difficile*	96.4 91.0
MPRL 4849	CBA	24 h	*C. difficile*	89.0 89.0
MPRL 4850	CBA	24 h	*C. difficile*	96.4 No spectrum
MPRL 4851	CBA	24 h	*C. difficile*	91.0 89.0
MPRL 4852	CBA	24 h	*C. difficile*	99.9 99.9
MPRL 4853	CBA	24 h	*C. difficile*	99.9 95.0
MPRL 4854	CBA	24 h	*C. difficile*	96.4 89.0
MPRL 4856	CBA	24 h	*C. difficile*	91.0 96.4
MPRL 4857	CBA	24 h	*Staphylococcus warneri*	
MPRL 554	CBA	24 h	*C. difficile*	89.0 91.0
MPRL 560	CBA	24 h	*C. difficile*	96.4 91.0
MPRL 586	CBA	24 h	*C. difficile*	99.9 95.0
MPRL 588	CBA	24 h	*C. difficile*	96.4 89.0
MPRL 589	CBA	24 h	*C. difficile*	91.0 96.4
MPRL 591	CBA	24 h	*C. difficile*	91.0 99.9
MPRL 613	CBA	24 h	*C. difficile*	89.0 91.0

SARAMIS™ identification of 43 (n= 43) isolates subcultured on CBA for 48 h analysed in duplicates. MPRL 13366 was identified as *P. acnes* on one read which could be due to the contamination of the sample. Isolate MPRL 307 was not identified from the 24 h culture as *C. difficile* was confirmed from the 48 h run.

Strain ID	Culture media	Incubation time	SARAMIS ID	Confidence level (%)
MPRL 002	CBA	48 h	*C. difficile*	96.4 99.9
MPRL 059	CBA	48 h	*C. difficile*	96.4 99.9
MPRL 1037	CBA	48 h	*C. difficile*	91.0 81.0
MPRL 13366	CBA	48 h	*P.acnes* *C. difficile*	90.0 91.0
MPRL 199	CBA	48 h	*C. difficile*	89.0 91.0
MPRL 223	CBA	48 h	*C. difficile*	96.4 96.4
MPRL 2282	CBA	48 h	No ID	No spectrum No spectrum
MPRL 2783	CBA	48 h	*C. difficile*	91.0 No spectrum
MPRL 279	CBA	48 h	*C. difficile*	96.4 89.0
MPRL 282	CBA	48 h	No ID	No spectrum No spectrum
MPRL 296	CBA	48 h	*C. difficile*	96.4 81.0
MPRL 307	CBA	48 h	*C. difficile*	No spectrum 91.0
MPRL 308	CBA	48 h	*C. difficile*	96.4 91.0
MPRL 371	CBA	48 h	*C. difficile*	86.0 91.0
MPRL 372	CBA	48 h	*C. difficile*	96.4 91.0
MPRL 379	CBA	48 h	*C. difficile*	96.4 86.0
MPRL 382	CBA	48 h	*C. difficile*	81.0 96.4
MPRL 396	CBA	48 h	*C. difficile*	99.9 99.9
MPRL 402	CBA	48 h	*C. difficile*	No spectrum 76.8
MPRL 405	CBA	48 h	*C. difficile*	71.0 No spectrum
MPRL 406	CBA	48 h	*C. difficile*	96.4 99.9
MPRL 407	CBA	48 h	*C. difficile*	71.0 81.0

MPRL 418	CBA	48 h	*C. difficile*	96.4 99.9
MPRL 422	CBA	48 h	*C. difficile*	65.0 No spectrum
MPRL 4846	CBA	48 h	*C. difficile*	81.0 71.0
MPRL 4847	CBA	48 h	*C. difficile*	99.9 96.4
MPRL 4848	CBA	48 h	*C. difficile*	89.5 86.0
MPRL 4849	CBA	48 h	*C. difficile*	91.0 No spectrum
MPRL 4850	CBA	48 h	*C. difficile*	76.8 81.0
MPRL 4851	CBA	48 h	*C. difficile*	96.4 91.0
MPRL 4852	CBA	48 h	*C. difficile*	95.0 99.9
MPRL 616	CBA	48 h	*C. difficile*	No spectrum 81.0
MPRL 617	CBA	48 h	*C. difficile*	91.0 No spectrum
MPRL 629	CBA	48 h	*C. difficile*	99.9 99.9
MPRL 665	CBA	48 h	*C. difficile*	99.9 99.9
MPRL 678	CBA	48 h	*C. difficile*	99.9 99.9
MPRL 687	CBA	48 h	*C. difficile*	91.0 No spectrum
MPRL 712	CBA	48 h	*C. difficile*	96.4 89.0
MPRL 808	CBA	48 h	*C. difficile*	96.4 99.9
T	CBA	48 h	*C. difficile*	96.4 91.0
027 (L)	CBA	48 h	*C. difficile*	89.0 No spectrum
B1	CBA	48 h	No ID	No spectrum No spectrum
MPRL 369	CBA	48 h	*C. difficile*	89.0 81.0

SARAMIS™ identification of 10 (n= 10) isolates subcultured on FAA for 48 h and 9 (n= 9) isolates subcultured on NA for 48 h, analysed in duplicates.

Strain ID	Culture media	Incubation time	SARAMIS ID	Confidence level (%)
MPRL 002	FAA	48 h	*C. difficile*	99.9 99.9
MPRL 4849	FAA	48 h	*C. difficile*	99.9 99.9
MPRL 296	FAA	48 h	*C. difficile*	99.9 96.4
027	FAA	48 h	*C. difficile*	96.4 91.0
MPRL 617	FAA	48 h	*C. difficile*	99.9 99.9
B1	FAA	48 h	*C. difficile*	96.4 91.0
MPRL 371	FAA	48 h	*C. difficile*	99.9 96.4
MPRL 2783	FAA	48 h	*C. difficile*	99.9 99.9
MPRL 1037	FAA	48 h	*C. difficile*	No spectrum 91.0
MPRL 223	FAA	48 h	*C. difficile*	91.0 81.0
027	NA	48 h	*C. difficile*	96.4 99.9
MPRL 4849	NA	48 h	No ID	No spectrum
MPRL 223	NA	48 h	*C. difficile*	99.9 96.4
MPRL 296	NA	48 h	*C. difficile*	99.9 99.9
MPRL 002	NA	48 h	*C. difficile*	96.4 99.9
MPRL 1037	NA	48 h	No ID	No spectrum
MPRL 2783	NA	48 h	*C. difficile*	99.9 96.4
B1	NA	48 h	*C. difficile*	96.4 91.0
MPRL 617	NA	48 h	No ID	No spectrum

Appendix IV

List of the forward and the reverse primers used in the VNTR study of *C. difficile* for all 47 loci.

Locus	Location	Forward primer sequence (label 5'-3')	Reverse primer sequence (label 5'-3')
CD1	755721-755950	FAM-TTAATTGAGGGAGAATGTTAAA	AAATACTTTTCCCACTTTCATAA
CD2	3688632-3688751	FAM-CTTAATACTAAACTAACTCTAACCAGTAA	TTATATTTTATGGGCATGTTAAA
CD3	3239736-3239835	NED-GTTTAGAATCTACAGCATTATTTGA	ATTGGAATTGAATGTAACAAAA
CD4	167124-167172	VIC-TGGAGCTATGGAAATTGATAA	CAAATACATCTTGCATTAATTCTT
CD5	1954913-1954939	PET-TTTTTGAAACTGAACCAACATA	ACAAAAGACTGTGCAAATATACTAA
CD6	664660-664705	PET-TGTATGAAGCAAGCTTTTTATT	AATCCAGCAATCTAATAATCCA
CD7	4116072-4116109	NED-GTTTTGAGGAAACAAACCTATC	GATGAGGAAATAGAAGAGTTCAA
CD8	692929-693015	VIC-AATTTTAAGTTAACGTTTTTCTACAT	AGCCATTTTTATCAATCCTTTCTAT
CD9	771338-771422	FAM-GTAGAAGGGGCAAATAATGAG	CCTTCTGGCTTCCTTGTAATA
CD10	677132-677386	PET-GGTGCACATGCTGGTCCTG	AACGCATTAAATTTCACTCCTCATAC
CD11	2168250-2168268	FAM-AGACAATGCACATATTCAACTCAC	GTAGTTTCATCAATTCTTACATCTAAT
CD12	804519-804537	VIC-CTTATCATAATTTCCAGACACAG	AAGTACAAAAAGGTAAAATTTGAAGG
CD13	301056-301074	NED-GTTATCTTCTGGGGTTAGAATC	TTGGTCTAAACCTTTTATTTGAGC
CD14	788448-788465	PET-GACATGTTAGACGAAGAAGTTCAA	TGATGCCTCCATATCACCAA
CD15	800026-800043	VIC-AAATAAACTTATTATAGGCAATCAAACA	ATTTTATCATCTAACTCACAAGTCAA
CD16	484804-484820	NED-AAGAAGAATGTGTTTAATATTGAGAGC	TCTTTTTAAGACTTACTGCATTCTTG
CD17	3254341-3254357	FAM-CATTCCAAAAGTATCTTTAGTTGC	GGTGATGGAGATTATGTTGATTTTT
CD18	3014724-3014741	NED-GGTGAGACGAAGATTAAAATACC	TTGTTCTAGTACCAAATAATCCGTTTT
CD19	2167938-2167955	VIC-GCCGAATTAAGAATAAAGTTGGTG	TTGGCAAAATATTTGTTGGATG
CD20	2169070-2169086	NED-GTGGAAGATATTGAGGCTGA	TCGGAAGTTCTGCTTCCTTT
CD21	2169418-2169435	PET -AATGCAAATTTGGAAGAAATGA	CCAGCATAATACCAGCCAAAA
CD22	804623-804645	VIC-GCACCTCATCACCATCTTCA	TCAAAATGAAAGACGACGAAAA
CD23	804652-804699	NED-CCTTTCTTCTCCTCTTCTTCTT	GGTCATTTCTAACCAAACATCAGT
CD24	478055-478070	FAM-CCACAGATAAAAATAAGAGGAAT	CTTGTCCTCTATCAAACCAAGCTAC
CD25	543601-543619	NED-CTTTCTGATTTTTCTTTTGTGTG	AGAACCCTACATAGGAACAACTTC
CD26	3245854-3245874	PET-CATTTGCATCTGCTCTTTTTAT	GGAAATAAGCTTGATACTACTCAAAAA

CD27	3246180-3246195	FAM-ATTTTTGCATTTGTCTCACTTC	CAGAGAATGGAACTGAAGTAGCTG
CD28	3246752-3246770	NED-TCCAGAACCCGTTAATGTATC	TTACGAAAGAATTAGGAGGTAAGAA
CD29	789540-789558	FAM-GTAGAGGAGACTTATCCTGGAA	GTGCAAATCAATATGAAGTTAGAATAA
CD30	794475-794493	VIC-GTATTATTTTGATCCTGATACAGCTC	TTGCTACTTTTCTGATTCCTCCA
CD31	3253605-3253625	VIC-GCAGTTTCTATTCTATCTTCTCCTG	GCAAATGATACAATAGCTAGTCAAGA
CD32	252622-252644	PET-CAATACTTTTAAGAAAAGGAATACAA	TAGCTTTTTACCTTGTTGAACTATTTG
CD33	800719-800741	FAM-GTTGGTAAAACTAATCTTGGATATG	AAAATAAAGTCACTTCCTTCTGTAGAC
CD34	799700-799717	FAM-CATTAAAACCAGTTTATGAAGACAC	CTGAAAAATCTATTGTTTGATTGCCTA
CD35	797131-797146	NED-AATTTTTAAACCAACACCTTAACC	TATTAAAGTCTACCCCATTGTCTTC
CD36	881618-881637	VIC-GAGCTGCTGCACTAGTAGG	GCCTTTTCCATTGTTTCTGC
CD37	881702-881718	NED-GGGCTTCCTAAAGAAACAG	TCACAACTCCCCTACAAGCA
CD38	883903-883918	FAM-GAAATGGCTCCTCCAAAAG	CATGTCCGTACCTGCCATTA
CD39	884179-884194	NED-TGAAATGGCTCCTCCAAAA	GGAAAAATTTAGCCCCTGCT
CD40	884601-884612	VIC-GGCAGGGGAACTTATCAAA	TCTGCACCAAGCATAGAAGC
CD41	886202-886220	FAM-ATGCTGATGCCAATGCTAA	GTCATACCACCAGGGAATGC
CD42	923152-923190	FAM-CTGAAACAGCACTTGACATTGAA	CCCTTTTTATTCATAATAATCATTCC
CD43	623650-623691	PET-AAGTGTGCTAAGTGTTGTTTTGC	TGATATAAAAAGACAAAAAGGGCAAT
CD44	771250-771319	NED-TCATTGAAGACAGAGATACAAAAA	GAAGCCTTCTGGCTTCCTTGT
CD45	3319803-3320164	PET-TTTCATAAAAGATTCCTTTCCTGT	AAGAGCATATATGGGAGTTTCTGTT
CD46	3753183-3753574	VIC-TCAGAAAATGCACCTTTAAATC	CCATAATAAGATTGCAGAGGTATAAGA
CD47	1099662-1099714	NED-GCAGGAACACCAATGGAGA	CCACAACCACCGTAACCTTT

Appendix V

Single-plex VNTR-PCR and multiplex VNTR-PCR reactions carried out for all 47 loci in the following combinations.

Loci	Single-plex/Multiplex
CD1	Single-plex PCR reactions
CD2	
CD3	
CD4	
CD5	
CD6	
CD7	
CD8	
CD9	
CD10	
CD11	
CD12	
CD13	
CD14	
CD15	
CD16	
CD17	
CD18	
CD29 CD19 CD21 CD35	Multiplex 1
CD28 CD26 CD22 CD33	Multiplex 2
CD27 CD20 CD30	Multiplex 3
CD31 CD34 CD32 CD23	Multiplex 4
CD24 CD25	Multiplex 5
CD36 CD39 CD41	Multiplex 6
CD37 CD38 CD40	Multiplex 7
CD42 CD43 CD44	Multiplex 8
CD45 CD46 CD47	Multiplex 9

Appendix VI

Specific proteins detected for isolates MW2 (MRSA) and 476 (MSSA)

Protein specific for MRSA	Protein specific for MSSA
hypothetical protein SAV2474	PREDICTED: similar to lambda-immunoglobulin [Equus caballus]
30S ribosomal protein S9	F0F1 ATP synthase subunit alpha
50S ribosomal protein L14	50S ribosomal protein L1
50S ribosomal protein L6	major tail protein
hypothetical protein SAV0378	hypothetical protein MW2433
30S ribosomal protein S4	ABC transporter ATP-binding protein
cytidylate kinase	triosephosphate isomerase
NAD-specific glutamate dehydrogenase	putative heme peroxidase
UDP-N-acetylmuramyl tripeptide synthetase-like	cell shape determinant mreC
S-adenosylmethionine synthetase	PREDICTED: similar to IGL@ protein [Equus caballus]
glutamate-1-semialdehyde aminotransferase	PREDICTED: similar to keratin 1 isoform 7 [Macaca mulatta]
GTP-binding protein EngA	F0F1 ATP synthase subunit beta
phosphoesterase	immunoglobulin gamma 7 heavy chain [Equus caballus]
translation-associated GTPase	6-phosphogluconate dehydrogenase
arginine deiminase	transcription elongation factor NusA
autolysin, N-acetylmuramyl-L-alanine amidase and endo-b-N-acetylglucosaminidas	PREDICTED: similar to Keratin, type I cytoskeletal 14
RecName: Full=Serum albumin; AltName: Full=BSA	30S ribosomal protein S1
RecName: Full=Chaperone protein dnaK; AltName: Full=Heat shock protein 70	immunoglobulin gamma 1 heavy chain constant region [Equus caballus]
alpha 2 globin [Equus caballus]	F0F1 ATP synthase subunit alpha
thimet oligopeptidase-like protein	
pyruvate oxidase	
hypothetical protein MW0374	
aspartyl-tRNA synthetase	
1-pyrroline-5-carboxylate dehydrogenase	
DNA gyrase, B subunit	
ATP-dependent Clp protease, ATP-binding subunit ClpC	