Genetic diversity of predominant nosocomial meticillin-resistant Staphylococcus aureus lineages in the United Kingdom

Thesis

Genetic Diversity of Predominant Nosocomial Meticillin-Resistant *Staphylococcus aureus* Lineages in the United Kingdom

Darshana Desai, BSc (Hons)

Molecular Microbiology

Submitted August 2013 to the Open University for the Degree of Doctor of Philosophy

Affiliated Research Centre: Public Health England

Date of Submission: 15 August 2013
Date of Award: 11 January 2014

NOTIFICATION OF REDACTION

THESIS TITLE:
Genetic diversity of predominant nosocomial meticillin-resistant Staphylococcus aureus lineages in the United Kingdom

AUTHOR:
Darshana Desai

YEAR:
2013

CLASSMARK:
616.920421 DES

The following pages/sections have been redacted from this thesis:

Page No.	Item/section redacted
346 - End	Publications

Contents

List of tables

List of figures

Acknowledgements

I confirm none of the material presented in this thesis has been previously submitted for any qualification to any university or institution. A proportion of the data has been published in the Journal of Medical Microbiology in 2012 and can be found at the end of this thesis.

Firstly I would like to thank Dr Russell Hope for providing the British Society for Antimicrobial Chemotherapy collection of isolates and Tom Nichols at Public Health England who performed statistical analysis of this collection.

I would especially like to thank all members of the Genomic Services Unit, for providing me with constant advice and support throughout the years.

To my supervisors Dr Meeta Desai, Dr Matthew Ellington and Dr Catherine Arnold, I will always be grateful for the patience, support and encouragement you have provided me throughout the years, thank you for all of your advice.

A special thank you goes to Millie and Hinal, I cannot tell you enough how your constant encouragement has got me through my write-up. Finally to my family who have supported me in every way so I can finish this thesis, Krina, Shivani and Ravi thank you for everything you have done to make writing up that little bit easier. This thesis is dedicated to my parents, who have always supported me and to whom I owe my every success.

Abstract

Meticillin-resistant *Staphylococcus aureus* (MRSA) is one of the major human pathogens and is a common cause of hospital-associated infection in immunocompromised patients and individuals with open wounds. However, infections affecting the healthy and young have increased in the past few decades and are a rising clinical concern. Understanding the genetic diversity of *S. aureus* can therefore result in a better understanding of the pathogenesis, spread and evolution of *S. aureus*, which in turn can lead to the development of effective infection control measures. In the present study, molecular methods were utilised to investigate differences in the distribution and variability of bacterial genes present between successful and unsuccessful lineages of MRSA. Lineages were assigned based on multilocus sequence typing (MLST) clonal complexes and successful lineages were defined as those lineages that have given rise to multiple epidemic MRSA clones. In addition to well-established molecular methods such as MLST and staphylococcal cassette chromosome *mec* typing, a novel fluorescent amplified fragment length polymorphism assay was developed to identify regions of genetic heterogeneity between and within lineages. These differences were attributed to lineage-specific sequence variation and differences in the distribution of mobile genetic elements (MGEs) between lineages. These genetic differences were investigated in relation to functional differences within the bacterium, to explore the role of cell functions encoded by these regions in the emergence and prevalence (or spread) of dominant genetic lineages. This study indicates a combination of factors play a role in the success of major MRSA lineages. The genetic diversity amongst core regions and MGEs appear to alter antimicrobial resistance and virulence factors that are vital to their success, as they enable the rapid adaption of lineages to their environment.

Abbreviations

AF	Amplified fragment
AFLP	Amplified fragment length polymorphism
AGA	N-acetyl glucosamine
agr	Accessory gene regulator
AMA	N-acetyl muramic acid
ATP	Adenosine triphosphate
attB	Bacterial chromosomal attachment site
BLAST	Basic Local Alignment Search Tool
bp	Base pair
BSAC	British Society for Antimicrobial Chemotherapy
BTS	Bacterial test standard
CA	Community-associated
CA-MRSA	Community-associated meticillin-resistant *Staphylococcus aureus*
CBA	Columbia agar supplemented with horse blood
CC	Clonal complex
CI	Confidence interval
Clf	Clumping factor
CMPHL	Clinical Microbiology and Public Health Laboratory
CoNS	Coagulase-negative staphylococci
D	Discriminatory power
ddNTP	Di-deoxynucleotide
DLV	Double-locus-variant
dNTP	Deoxynucleotide

dru	Direct repeat unit
Eap	Extracellular adherence protein
EARSS	European Antimicrobial Resistance Surveillance System
Efb	Extracellular fibrinogen-binding protein
EMRSA	Epidemic meticillin-resistant *Staphylococcus aureus*
FAFLP	Fluorescent amplified fragment length polymorphism
FAM	5'-carboxyfluorescein
GI	Genomic island
GISA	Glycopeptide-intermediate *S. aureus*
GRSA	Glycopeptide-resistant *S. aureus*
HA	Hospital-associated
HA-MRSA	Hospital-associated meticillin-resistant *Staphylococcus aureus*
HCCA	α-cyano-4-hydroxycinnamic acid
HGT	Horizontal gene transfer
hVISA	Heterogeneous vancomycin-intermediate *Staphylococcus aureus*
ICU	Intensive care unit
IS	Insertion sequence
IS-FAFLP	Insertion sequence targeted fluorescent amplified fragment length polymorphism
IWG-SCC	International Working Group on the classification of Staphylococcal Cassette Chromosome elements
J region	Joining region
kb	Kilobase pair
LIZ	The chemical name is unknown as this is a proprietary fluorescent dye from Life Technologies, UK

MALDI-TOF MS	Matrix-assisted laser desorption ionization-time-of-flight mass spectrometry
Mb	Megabase pair
MGE	Mobile genetic element
MHC	Major histocompatibility complex
MIC	Minimum inhibitory concentration
MLEE	Multilocus enzyme electrophoresis
MLST	Multilocus sequence typing
MRSA	Meticillin-resistant *Staphylococcus aureus*
MSCRAMM	Microbial surface components recognising adhesive matrix molecules
MSSA	Meticillin-susceptible *Staphylococcus aureus*
NCBI	National Center for Biotechnology Information
ND	No data
NED	1',2'-benzo-4'-fluoro-7'-chloro-6-carboxy-4,7-dichlorofluorescein
NGS	Next generation sequencing
NT	Non-typeable
ORF	Open reading frame
PBP	Penicillin-binding protein
PCR	Polymerase chain reaction
PET	4-(N, N-Dimethylaminoethylene) amino-N-allyl-1,8-naphthalimide
PFGE	Pulsed-field gel electrophoresis
PHE	Public Health England
PPi	Pyrophosphate
PVL	Panton-Valentine leukocidin
RFU	Relative fluorescent units

R-M	Restriction-Modification
rRNA	Ribosomal ribonucleic acid
SaPI	*Staphylococcus aureus* pathogenicity island
SCC*mec*	Staphylococcal cassette chromosome *mec*
SCV	Small-colony variant
SE	Staphylococcal enterotoxin
SEl	Staphylococcal enterotoxin-like protein
SIRU	Staphylococcal interspersed repeat unit
SLV	Single-locus-variant
SNP	Single-nucleotide polymorphism
spa	Staphylococcal protein A gene
SRL	Staphylococcal Reference Laboratory
ST	Sequence type
TFA	Trifluoroacetic acid
TSST-1	Toxic shock syndrome toxin-1
UPGMA	Unweighted pair group method with arithmetic mean
VIC	2'-chloro-7'-phenyl-1,4-dichloro-6-carboxyfluorescein
VISA	Vancomycin-intermediate *S. aureus*
VNTR	Variable-number tandem-repeat
VRSA	Vancomycin-resistant *Staphylococcus aureus*
WGM	Whole-genome mapping
WGS	Whole-genome sequence

CHAPTER 1

1. Introduction

1.1. The genus *Staphylococcus*

1.1.1. The history of *Staphylococcus*

Alexander Ogston was the first to detail the characteristic grape-like clusters of staphylococci in 1882. He named the genus *Staphylococcus*, based on the Greek word 'staphyle' meaning bunch of grapes (Ogston, 1880). Subsequent cultivation by Rosenbach provided evidence that the microorganism could cause wound infections, which in the pre-antimicrobial era often endangered life (Rosenbach, 1884). Rosenbach distinguished two strains of *Staphylococcus* which he named on the basis of colony pigmentation: *Staphylococcus aureus* based on the Latin word 'aurum' for gold and *Staphylococcus albus* based on the Latin word 'albus' for white.

Staphylococci were initially grouped into the family Micrococcaceae with micrococci into the genus *Micrococcus* (Macdonald and Smith, 1981). However, studies based either on single biochemical characteristics or on multiple characteristics using numerical taxonomy displayed both genera were distinct (Evans *et al.*, 1955; Hill, 1959). Analysis of the percentage G+C content further supported this distinction. *Staphylococcus* displayed a low G+C content of 31-37 % whilst *Micrococcus* exhibited a higher G+C content of 69-75 % (Silvestri and Hill, 1965). 16S ribosomal RNA (rRNA) gene sequence analysis also identified staphylococci were distinct from *Micrococcus* and the other genus in the family, *Planococcus* (Stackebrandt and Woese, 1979). Kloos *et al.* (1998) have subsequently shown that *Staphylococcus* is closely related to the more recently described genus *Macrococcus*. Since then, the genus *Staphylococcus* has been assigned to the family Staphylococcaceae along with *Macrococcus*. This genus is monophyletic (descended from a single common ancestor). Over

40 species and 20 subspecies within the genus *Staphylococcus* have been described to date (de Vos *et al.*, 2009).

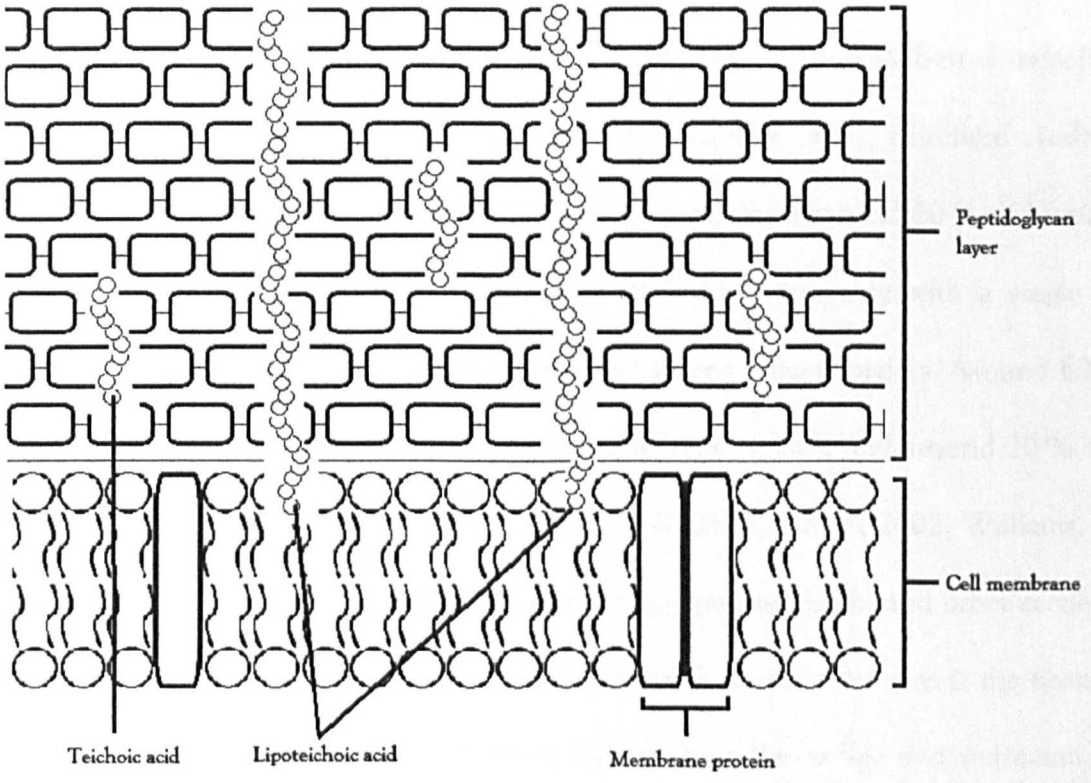
1.1.2. Characteristics of *Staphylococcus aureus*

Staphylococci are Gram-positive bacteria which commonly form part of the human and animal microbiota. Species such as *S. aureus* and *S. epidermidis* are colonisers of human skin and mucosa, and are a common cause of human infections. *S. aureus* exists in two forms: planktonic cells (free living cells) or within biofilms (aggregation of cells to a surface). The cells have a diameter of 0.5–1.5 μm and grow in characteristic grape-like clusters as they divide along three planes. The colonies can appear golden yellow due to the synthesis of a carotenoid pigment but are more commonly cream, white or non-pigmented. It is a non-spore forming, non-motile and facultative anaerobic bacterium. Cells have an optimum growth temperature of between 30 to 37°C and they can tolerate a pH ranging from 4 to 10. *S. aureus* is catalase- and coagulase-positive; the latter enables clotting of blood plasma and enables it to be distinguished from the approximately 40 coagulase-negative staphylococci (CoNS) species such as *S. epidermidis* (Harris *et al.*, 2002). *S. aureus* can utilise various carbon sources and are capable of growth *in vitro* on a variety of media. When grown on blood agar, some colonies show a zone of β-haemolysis due to the production of haemolysin, which results in the lysis of erythrocytes (Singleton, 2004).

S. aureus possesses a plasma membrane surrounded by a cell wall which is 20-50 nm thick. The cell wall maintains the cell shape and protects the cell against osmotic forces. Approximately 40 % of the cell wall is composed of peptidoglycan also known as murein (see penicillin-resistant *S. aureus*: section 1.4.2) and other polymers such as teichoic acid (Figure 1) (Shockman & Barrett, 1983). The peptidoglycan and teichoic acid compose 90 % of the

cell wall. The remaining component is made up of surface proteins and exoproteins which may play a role in virulence (Harris *et al.*, 2002). In addition, around 90 % of *S. aureus* isolates produce capsular polysaccharides which may enhance the virulence of the organism (Thakker *et al.*, 1998).

Figure 1. Cross-sectional view of a Gram-positive bacterium cell wall



The cell membrane depicts membrane proteins and a lipid bilayer, the latter anchors lipoteichoic acid. Lipoteichoic acid and teichoic acid are found within the peptidoglycan layer.

1.2. Clinical significance of *S. aureus*

1.2.1. Colonisation

S. aureus causes infections in humans and a variety of animal hosts. *S. aureus* carriage has been described in reptiles, birds and mammals, the latter of which include livestock and domestic pets. There is no consensus between studies for the rate of carriage in canines and felines but prevalence rates range from 0 to 4 % (Weese, 2010). Risk factors include open wounds, previous antimicrobial treatment and contact with colonised individuals (Thompson *et al.*, 1982). *S. aureus* is ubiquitous in nature and around 20 % of humans are thought to be permanent carriers and are colonised continuously with a single strain. However, a turnover of strains has been reported in permanent carriers. Around 60 % are thought to be intermittent carriers where the strains are variable and around 20 % are not carriers (Sakwinska *et al.*, 2009; Singleton, 2004; Whitt & Salyers, 2002; Williams, 1963). Colonisation has recently been reclassified into two types, persistent and other carriers (van Belkum *et al.*, 2009). The most common reservoir of *S. aureus* in humans is the nostrils but they can also be found in areas of moist skin such as the axillae and perineum. Nasal colonisation has been associated as a risk factor for *S. aureus* infections in the immunocompromised, the elderly and the very young, as well as those recovering from open surgery (Frank *et al.*, 1999). In infections such as bacteraemia, the majority of patients are infected by the same strain they are colonised with (von Eiff *et al.*, 2001; Wertheim *et al.*, 2004). Similarly, a higher proportion of children are persistently colonised as compared with adults. The causes of this are as yet unknown, although bacterial cell surface and adhesion proteins are thought to play a role in colonisation (Armstrong-Esther & Smith, 1976; Kluytmans *et al.*, 1997; Nouwen *et al.*, 2004).

1.2.2. Types of disease

Staphylococci are a common cause of skin and soft tissue infections and bacteraemia in humans. In the broad term bacteraemia refers to the any presence of bacteria in the blood. However, overtime it has become used to refer to the presence of bacteria in the blood which result in clinical symptoms. This latter definition has been utilised in this thesis here after. Symptoms associated with infection vary depending on the site of infection. Most common superficial infections develop in hair follicles and present as boils (or furuncles) which are pus-filled cavities. Deeper penetration can result in internal abscess formation (Ogston, 1880) and penetration of the organism into the bloodstream which can seed infections at other body sites and lead to arthritis, meningitis and pneumonia. This is common in patients with predispositions such as existing illness and old age. Intravenous lines and open wounds also increase chances of bacteraemia (Fluit & Schmitz, 2003).

S. aureus can encode acquired toxins which act as superantigens and cause toxigenic diseases. Clinical symptoms that are associated with toxigenic staphylococcal disease include food poisoning, staphylococcal scalded skin syndrome, toxic shock syndrome and necrotising fasciitis, amongst others. Toxic shock syndrome is caused by the invasion of a toxin into the bloodstream which can result in toxic shock and multi-organ failure (Lappin & Ferguson, 2009). Staphylococcal food poisoning is caused by consuming foods products that contain one or more staphylococcal emetic toxins. Toxins cause symptoms ranging from nausea to severe dehydration and changes in blood pressure. Infections also manifest as the contagious condition, bullous impetigo, which is found to spread rapidly between infants and close contacts. This is characterised by the formation of blisters containing large numbers of bacteria in the top layers of skin (Fluit & Schmitz, 2003).

1.2.3. Burden of disease

Meticillin-resistant *S. aureus* (MRSA) (see MRSA: section 1.4.3) is among the most common antimicrobial-resistant pathogen identified among hospitals around the world, although the incidence of MRSA varies between countries (European Antimicrobial Resistance Surveillance System annual report 2008). The UK continued to show a decreasing trend in prevalence down to 31 % in 2008. The European Antimicrobial Resistance Surveillance Network attributed 1 % of *S. aureus* infections in Northern European countries to MRSA and 50 % to MRSA in Southern European countries (European Antimicrobial Resistance Surveillance System annual report 2008). This included the first case of MRSA reported in Iceland between 1999 and 2008. Countries such as Belgium, Poland and France reported a continued decrease whilst others such as Portugal and Switzerland reported an increase. In 2005, over half of hospitalisations due to *S. aureus* infections in the USA were due to MRSA (Klein *et al.*, 2007). During the period 2005 to 2008, a 9.4 % decrease in hospital-associated (HA) infections and a 5.7 % decrease in community-associated (CA) infections was observed in the USA (Kallen *et al.*, 2010).

Throughout 2001 to 2006, the proportion of *S. aureus* bacteraemia attributable to MRSA was around 42 % in the UK and Ireland and these strains often displayed multiple resistances (Hope *et al.*, 2008). Although the proportion of MRSA bacteraemia in 2009 and 2010 dropped to 23 % and 19 % respectively, the rate of decline has slowed since 2006 (Martin *et al.*, 2011). Despite the overall decrease, the proportion of bacteraemia whose source of infection was attributed to skin and soft tissue infections has increased (Pearson *et al.*, 2009; Wilson *et al.*, 2011). A study of nosocomial bacteraemia during the period 1996 to 2006 from more than 16 countries suggested that 60.2 % of infections were caused by

epidemic MRSA (EMRSA)-15 and 35.4 % by EMRSA-16 (Johnson *et al.*, 2001). These two strains have continuously accounted for the majority of MRSA bacteraemia throughout 2001 to 2007 (Ellington *et al.*, 2010), despite an overall reduction of 56 % in MRSA bacteraemia between 2004 and 2008 (Johnson *et al.*, 2012). In the last decade, a decline in EMRSA-16 in the UK was observed which was compensated by a proportionate increase in the prevalence of EMRSA-15. EMRSA-16 declined dramatically from 21.4 to 9 % during the period 2001 to 2007 whilst the proportion attributable to EMRSA-15 increased correspondingly (Ellington *et al.*, 2010). EMRSA-16 may however have been in natural decline before the implementation of enhanced infection control measures (Rao *et al.*, 2011; Wyllie *et al.*, 2011).

1.3. Virulence and pathogenicity factors

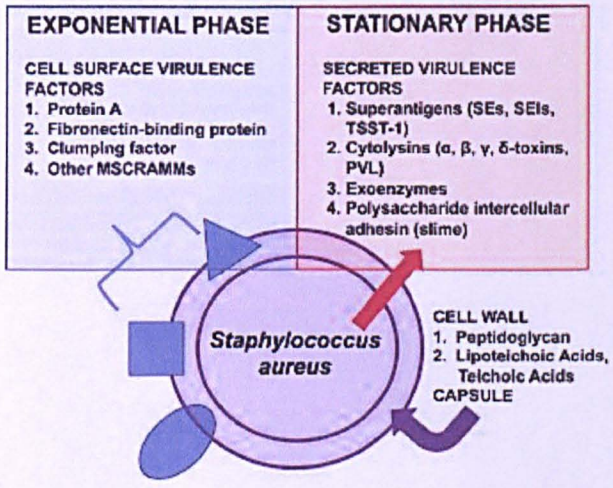
Virulence factors assist the organism to adhere to host tissue, evade the host immune system and to disseminate and invade tissue (Projan & Novick, 1997). Virulence factors also affect the degree of pathogenicity, which is the ability of an organism to cause disease (van Belkum *et al.*, 2007). The production of virulence factors varies between growth stages (Figure 2) (Schlievert *et al.*, 2010). The presence of *S. aureus* infections proves that this organism can become established and transmit affectively between hosts. The production of virulence and pathogenicity factors often found on mobile genetic elements (MGEs) such as bacteriophages and *S. aureus* pathogenicity islands (SaPIs) are also contributing factors.

1.3.1. Factors which aid tissue adhesion

Fibronectin-binding proteins enable adhesion of bacterial cells to fibronectin (plasma protein) (Jönsson *et al.*, 1991) and subsequently bind to the fibronectin receptor (integrin

α5β1). This facilitates the uptake of bacteria into host cells (Schwarz-Linek *et al.*, 2003; Sinha *et al.*, 1999). Mutations within the proteins reduce the binding capacity of *S. aureus* to fibrinogen and its ability to invade host cells (Grundmeier *et al.*, 2004). Clumping factors ClfA and ClfB enable adhesion of *S. aureus* to fibronectin and cytokeratin, which results in the aggregation of platelets and protects the bacteria against phagocytosis by host leukocytes (Bayer *et al.*, 1995; Palmqvist *et al.*, 2004). The above proteins are examples of microbial surface components recognising adhesive matrix molecules (MSCRAMM) (Holden *et al.*, 2008).

Figure 2. *S. aureus* virulence factors produced during different bacterial growth stages



Cell surface virulence factors are produced during the exponential growth stage whilst secreted factors are produced during the stationary phase (obtained from Schlievert *et al.*, 2010). MSCRAMM, Microbial surface components recognising adhesive matrix molecule; PVL, Panton-Valentine leucocidin; SE, staphylococcal enterotoxin; SEl, staphylococcal enterotoxin-like protein; TSST-1, toxic shock syndrome toxin-1.

The extracellular adherence protein (Eap) and extracellular fibrinogen-binding protein (Efb) are *S. aureus* secreted proteins that facilitate their adhesion to fibrinogen (Palma *et al.*, 1999). Eap decreases the migration of neutrophils to the site of infection (Chavakis *et al.*, 2002) and prevents extravasation (movement from capillaries to tissue) of leukocytes (Harraghy *et al.*, 2003) whilst Efb prevents complement activation of T-cells (Lee *et al.*, 2004a; 2004b). Sortase A (enzyme anchoring surface adhesion proteins) mutants display a decrease in virulence, thus highlighting the importance of cell adhesion to the pathogenesis of *S. aureus* (Fluit & Schmitz, 2003).

1.3.2. Factors which aid immune system evasion

Staphylococcal protein A is anchored within the cell wall and prevents phagocytosis and initiates apoptosis (programmed cell death) of host leukocytes (Goodyear & Silverman, 2004). The chemotaxis inhibitory protein of *S. aureus* is a secreted protein that binds to the C5a receptor on leukocytes and prevents their induction (de Haas *et al.*, 2004; Wright *et al.*, 2007). Various *S. aureus* strains produce a capsule layer and the majority of strains causing infection have been shown to belong to capsule serotype 5 or 8. The capsule and other secreted molecules such as the staphylococcal complement inhibitor and staphylokinase protein prevent the binding of antibodies to the bacterial cell surface preventing recognition and phagocytosis by host leukocytes (Karakawa *et al.*, 1988; Rooijakkers *et al.*, 2005a; 2005b; van Wamel *et al.*, 2006). *S. aureus* produce proteases such as the cysteine protease staphopain B and the metalloprotease aureolysin (Massimi *et al.*, 2002) which degrade host tissue (Lowy, 2000) and plasma proteins (Potempa *et al.*, 1986; 1991).

S. aureus produces superantigens that are presented on the major histocompatibility complex class II (MHC II) molecules at the cell surface of antigen-presenting cells during an infection.

These MHC II molecules interact with receptors on T-leukocytes which generate an increase in cytokine production, a surge in immune cells and the inflammatory response. Superantigens can be encoded on MGEs such as bacteriophages and SaPIs (Betley & Mekalanos, 1985). Superantigens include the enterotoxins (types A to E, G to R and U) and the production of certain variants of these toxins cause staphylococcal food poisoning (Jarraud *et al.*, 2001; Letertre *et al.*, 2003; Omoe *et al.*, 2003; Orwin *et al.*, 2001; 2003). The *egc* toxin gene cluster, which encodes the toxins, is found in specific staphylococcal lineages (Fluit & Schmitz, 2003). Lineages are assigned based on multilocus sequence typing (MLST) clonal complexes. Enterotoxin production has been shown to differ between colonising and bacteraemia isolates (Peck *et al.*, 2009). Other *S. aureus* toxins such as toxic shock syndrome toxin-1 (TSST-1) which plays a role in the apoptosis of host cells, can lead to toxic shock syndrome (Hofer *et al.*, 1996; Kreiswirth, 1989; Lowy, 2000), whilst staphylococcal scalded-skin syndrome is associated with exfoliative toxin A (Ladhani, 2003).

1.3.3. Factors which aid tissue invasion

Cytotoxins secreted by *S. aureus* lyse host leukocytes and erythrocytes (Song *et al.*, 1996). Three forms of haemolysins: α, β and γ are produced by *S. aureus*. Alpha-haemolysin forms pores in the cell membrane of erythrocytes (Bhakdi & Tranum-Jensen, 1991; Bohach & Foster, 2000) whilst β-haemolysin degrades sphingomyelin in the cytoplasmic membrane to cause cell lysis (Bohach & Foster, 2000). The third form, γ-haemolysin, works synergistically with leukocidins such as Panton-Valentine leukocidin (PVL) to form pores in leukocytes (Jayasinghe & Bayley, 2005; Joubert *et al.*, 2006). Strains with mutations in the gene encoding γ-haemolysin have been associated with persistent bacteraemia (Fowler *et al.*, 2004) whilst PVL is thought to play a role in necrotising pneumonia (Labandeira-Rey *et al.*, 2007). PVL is a toxin composed of two subunits encoded by the genes *luk*S-PV and *luk*F-PV

(Rahman *et al.*, 1992). It is detected to a greater extent in clinical isolates as compared to colonising isolates (Chiu *et al.*, 2011). There has been a rise from 2 to 15 % in clinical isolates that are PVL-positive (Prevost *et al.*, 1995; Wannet *et al.*, 2005).

Invasive and carriage isolates have been shown to be genetically similar and occur in similar lineages (Lindsay *et al.*, 2006; Melles *et al.*, 2004). However, variation among virulence factors such as surface proteins is greater between lineages than within lineages. This variation is lineage-specific, although variation among protein domains is conserved across unrelated lineages (McCarthy & Lindsay, 2010). The absence of certain virulence factors in lineages associated with invasive strains questions the role of individual factors in virulence and suggests combinations of proteins are required to confer increased virulence. Despite this, virulence genes do differ between strains and some of these are important for a change from colonisation to an invasive pathogen (Peacock *et al.*, 2002). Whether virulence genes differ between MRSA and meticillin-susceptible *S. aureus* (MSSA) is controversial, although factors such as enterotoxins do differ between genetic backgrounds (Liu *et al.*, 2010; Peck *et al.*, 2009; Varshney *et al.*, 2009).

1.3.4. Survival mechanisms of *S. aureus*

S. aureus virulence factors are regulated by global regulators, which help the bacterium to adapt to its environment and survive. The accessory gene regulator (*agr*) locus regulates the transcription of the protein A (*spa*), β-haemolysin (*hlb*) and toxic shock syndrome toxin-1 (*tst*) genes (Recsei *et al.*, 1986). The *agr* system plays a role in quorum sensing, which is when bacteria detect the bacterial population density in their vicinity by signalling molecules that cause changes in gene expression of exoproteins and proteins associated with cell walls (Cassat *et al.*, 2006; Ji *et al.*, 1995). *agr* groups I, II and III, which are defined based on their

sequence, are present in each lineage and are part of the variable genes (Lindsay *et al.*, 2006; Peacock *et al.*, 2002). These groups are specific to genetic background and may play a role in colonisation and pathogenesis (Goerke *et al.*, 2003; Schwan *et al.*, 2003). The system is involved in the interaction of *S. aureus* with the host cell surface and plays a role in the induction of apoptosis of host cells (Schnaith *et al.*, 2007) and the shedding of extracellular proteins from host cells. Other regulators such as the staphylococcal accessory element system is important in the adhesion to host cells and in the induction of apoptosis of host cells (Liang *et al.*, 2006a; 2006b) whilst the staphylococcal accessory gene regulatory locus is involved in the regulation of biofilm formation (Cucarella *et al.*, 2001).

S. aureus can form biofilms, an aggregation of cells that have adhered to a surface which are associated with persistent infections as antimicrobials cannot penetrate the cells (Costerton *et al.*, 1999). Strains within biofilms are often resistant to multiple (≥2) antimicrobials (Kwon *et al.*, 2008). Cells in biofilms display differential gene expression patterns, a reduced growth rate and altered metabolic state (Beenken *et al.*, 2004; Stewart & Costerton, 2001). *S. aureus* also produces bacteriocins (also known as lantibiotics), small molecules encoded by the *bsa* genes, which kill other species. Bacteriocins are encoded on MGEs and chromosome and are variable between lineages. Such molecules confer a selective advantage in a mixed population or under conditions of selection due to environmental stress.

1.4. Antimicrobial resistance in *S. aureus*

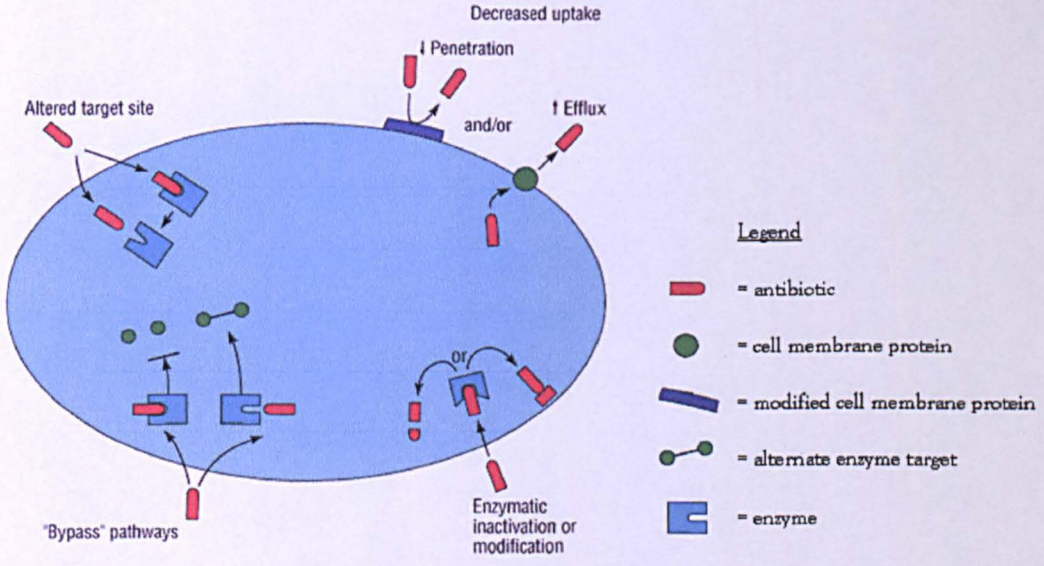
An antimicrobial is a substance of microbial or synthetic origin which kills or inhibits the growth of microbes. They have specific targets and are even active at low concentrations (Waksman, 1947). There are two types of antimicrobials: those displaying bactericidal

activity (kill bacteria) and those that are bacteriostatic (prevent growth). Antimicrobials are categorised as broad spectrum if they act on Gram-positive and Gram-negative bacteria or narrow spectrum if they are affective against a smaller group of bacteria.

1.4.1. Mechanisms of antimicrobial resistance

Various mechanisms confer antimicrobial resistance in bacteria (Figure 3). Some of these are always active, while others are induced in the presence of an antimicrobial (Singleton, 2004).

Figure 3. Mechanisms of antimicrobial resistance



The figure depicts four mechanisms of antimicrobial resistance. One mechanism involves the inactivation or modification of the antimicrobial molecules whilst another is the prevention of antimicrobial molecules penetrating into the cell or the removal of antimicrobial molecules via efflux pumps. Another mechanism involves modification of the target site that prevents the action of the antimicrobial. The cell can also 'bypass' the effect of the antimicrobial by production of an alternative target (obtained from Hawkey, 1998).

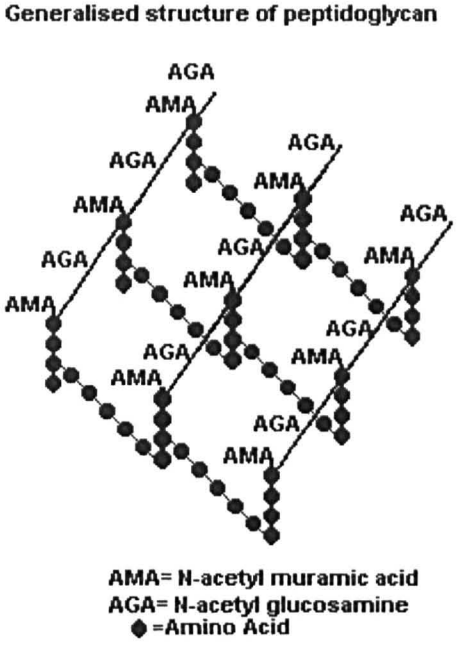
Two major mechanisms of resistance to β-lactam antimicrobials have been extensively studied in *S. aureus*. The first is the production of β-lactamases, which hydrolyse the β-lactam ring of the antimicrobial molecules. The second is the modification of the β-lactam targets known as penicillin-binding-proteins (PBPs) (Malouin & Bryan, 1986). Other mechanisms of resistance include the removal of antimicrobial molecules once they have penetrated the cell via efflux pumps embedded in the bacterial cell envelope (Figure 3). Resistance can also be conferred by changing the cell envelope composition, thus preventing the molecules from penetrating the bacterium or modification of the antimicrobial molecules (Wilke *et al.*, 2005). Further mechanisms involve production of an alternative target or the removal of the antimicrobial target.

1.4.2. Penicillin-resistant *S. aureus*

Penicillin was introduced clinically in 1941, however, its overuse may have driven or contributed to the selective pressure responsible for emergence of penicillin-resistant bacteria. Bacteria were identified prior to the introduction of penicillin which could produce proteins that conferred resistance to penicillin (Abraham & Chain, 1940). β-lactam molecules including penicillin contain a four member ring, one member of which is nitrogen. This class of antimicrobials prevent the synthesis of the cell envelope in growing cells. The cell wall is composed of many layers of peptidoglycan, a unique bacterial polymer comprising a glycan backbone made of alternating units of N-acetyl glucosamine and N-acetyl muramic acid, with short peptides cross-linking the N-acetyl muramic acid residues (Figure 4). The precise structure of the peptide cross-bridges can differ between bacterial species but always contain both D- and L-amino acids. Peptidoglycan is unique as it contains amino acids in the D-isomer form which help protect the cell from peptidases. Antimicrobial molecules saturate PBPs found in the cytoplasmic membrane by covalently binding to the

serine residue at the active site. PBPs catalyse the transglycosylation and transpeptidase reactions which form glycosidic and peptide bonds respectively of the peptide cross-bridges. The PBPs are therefore required in the synthesis of the peptidoglycan layer (Ghuysen, 1991; Malouin & Bryan, 1986). Meticillin thus results in a weakened peptidoglycan layer, predisposing the cell to lysis (Wilke *et al.*, 2005).

Figure 4. Structure of peptidoglycan

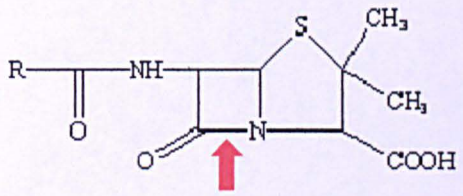


Peptidoglycan is a component of the bacterial cell wall. It is made of alternating units of N-acetyl glucosamine and N-acetyl muramic acid, with short peptides cross-linking the N-acetyl muramic acid residues (reproduced with permission from John Heritage, University of Leeds, UK).

The first mode of β-lactam resistance is conferred by the production of β-lactamases, which hydrolyse the amide bond in the β-lactam ring, thus inactivating antimicrobials such as penicillin (Figure 5) (Wilke *et al.*, 2005). In staphylococci, β-lactamases are encoded on

plasmids, which can transfer between strains thereby increasing the spread of resistance. β-lactamase production is encoded by the *blaZ* gene. Transcription of the gene is induced by the presence of β-lactam molecules and is regulated by the *blaI* repressor and *blaR1* receptor. β-lactamases are not affective against some β-lactam antimicrobials including meticillin (Singleton, 2004).

Figure 5. Structure of penicillin



The figure depicts the structure of penicillin. The red arrow indicates the β-lactamase target in the β-lactam ring.

1.4.3. Meticillin-resistant *S. aureus* (MRSA)

By the late 1950s, up to 80 % of *S. aureus* strains causing infections were resistant to penicillin. This increased the pressure to develop and discover novel antimicrobials. The semi-synthetic β-lactam antimicrobial meticillin was derived from penicillin and was introduced into Europe in 1959 for the treatment of infections due to penicillin-resistant *S. aureus*. However, the first MRSA were reported shortly after the introduction of meticillin (Jevons *et al.*, 1961). MRSA can express heterogeneous resistance, characteristic of strains from the 1980s or homogeneous resistance. Heterogeneous MRSA exhibit low or borderline minimum inhibitory concentrations (MICs) of meticillin, as low as 1.5 mg/L, whilst homogeneous MRSA exhibit higher MICs, >400 mg/L (Resende and Figueiredo, 1997). Heterogeneous MRSA often express resistance at low temperatures such as 30°C. However,

as cultures in current practice are incubated at 37°C, such strains may generally be identified as MSSA (Cookson, 2011).

The second mode of β-lactam resistance in *S. aureus* is conferred by a modified PBP. MRSA express four native PBPs in addition to a modified PBP2' or 2A. Although MRSA display reduced levels of peptidoglycan cross-linking, PBP2A is still able to catalyse the transpeptidation reaction for the cross-linking of the peptidoglycan layer and is therefore essential for cell wall assembly in the presence of β-lactams (Wyke *et al.*, 1982). PBP2A is encoded by the *mecA* gene found on the staphylococcal cassette chromosome *mec* (SCC*mec*; see MGEs: section 1.5.2) (Fluit & Schmitz, 2003). This protein (78 kDa) is only expressed in MRSA and has been modified and exhibits a lower affinity for meticillin. This reduced affinity lowers the rate at which the β-lactam molecules covalently bind to PBP2A. Cells therefore require higher concentrations of meticillin to saturate the PBP2A (Hartman & Tomasz, 1981; 1984).

The *mecA* gene is regulated by *mecR1* and *mecI* found on the SCC*mec*, although additional mechanisms are needed to regulate *mecA* (Oliveira & de Lencastre, 2011). The *blaR1* and *blaI* regulate the expression of the *blaZ* gene; purifying selection on the *bla* locus maintains its function even in MRSA. This correlates with the findings that *bla* regulatory genes can also regulate *mecA* (Milheiriço *et al.*, 2011). Apart from the *mecA* gene, the *aux* (auxiliary) and *fem* (factor essential for meticillin resistance) genes are essential for complete meticillin resistance (de Lencastre & Tomasz, 1994; de Lencastre *et al.*, 1999).

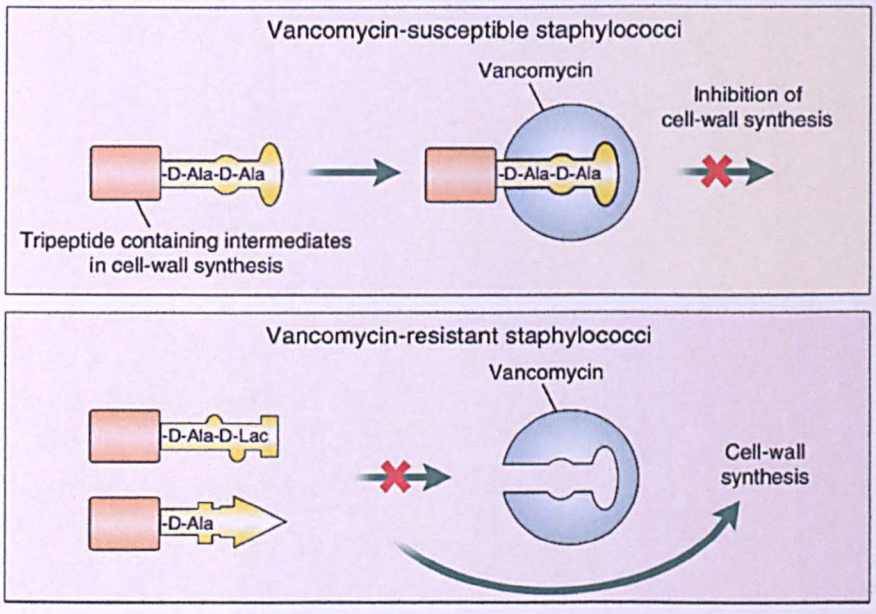
1.4.4. Vancomycin-resistant *S. aureus* (VRSA)

While glycopeptides such as vancomycin remain affective against the majority of *S. aureus*, where resistance occurs, it causes utmost concern due to the widespread use of the compounds as anti-MRSA agents. Glycopeptides bind to the end amino acids of peptidoglycan precursors, thus preventing the transpeptidase activity of PBPs and cross-linking of the peptidoglycan layer (Frère, 1995). The first *S. aureus* strains that displayed increased vancomycin MICs emerged in 1997 in Japan and were termed vancomycin-intermediate *S. aureus* (VISA) (Hiramatsu *et al.*, 1997). The mechanism that confers intermediate vancomycin resistance is still not understood (McAleese *et al.*, 2006). However, VISA have been demonstrated to have a thickened cell wall which prevents the antimicrobial molecules from penetrating further (Cui *et al.*, 2000). It has also been suggested that glycopeptides inhibit the transglycosylation reaction of PBPs (Ge *et al.*, 1999). On the contrary, heterogeneous vancomycin-intermediate *S aureus* (hVISA) give rise to some cells partially resistant to vancomycin (Hiramatsu *et al.*, 1997). CoNS are reported to act as a reservoir for vancomycin resistance (Nouwen *et al.*, 2005). Whilst such strains remain a concern, their true prevalence and impact on clinical care remain a matter of debate. There are concerns that these strains have been reported as being significantly associated with persistent bacteraemia and heart failure. Genotypes of these strains do not differ from strains not displaying heterogeneous vancomycin resistance.

In 2002, the first strains were reported that displayed high-level vancomycin resistance (MIC >256 mg/L). Fortunately these resistant strains have not spread widely. There have only been reports of sporadic incidents, including 11 vancomycin-resistant *S. aureus* (VRSA) cases in the USA in 2009 (Périchon & Courvalin, 2009). This resistance is thought to be conferred

by the *vanA* cluster found on the transposon Tn*1546*. Genes in this cluster cause structural changes in the peptidoglycan precursors which lowers the affinity of the vancomycin molecules to the precursors (Figure 6) (Bugg *et al.*, 1991). The *vanA* cluster may have been transferred from *Enterococcus faecalis* (Noble *et al.*, 1992). However, the plasmid harbouring Tn*1546* has varied between VRSA, indicating multiple insertions into *S. aureus*. Strains harbouring both the *mecA* and *vanA* gene have been reported and these can exhibit resistance to meticillin and vancomycin. However, such strains cannot grow in the presence of both antimicrobials as one resistance mechanism hinders the other (Severin *et al.*, 2004).

Figure 6. Mechanism of resistance to vancomycin



In the presence of vancomycin, the *vanA* operon enables the synthesis of a cell wall precursor ending in D-Ala-D-Lac dipeptide rather than D-Ala-D-Ala. The D-Ala-D-Lac dipeptide displays reduced affinity for vancomycin (obtained from Murray, 2000).

1.4.5. Evolution of antimicrobial resistance

Antimicrobial resistance mechanisms such as efflux pumps and enzymes that degrade or alter the antimicrobial may be encoded on the chromosome (intrinsic resistance). Conversely, resistance can arise via mutations in the genome or through the acquisition of MGEs via horizontal gene transfer (HGT; acquired resistance). In the latter case, these genes are thought to originate from environmental bacteria within which intrinsic resistance has been described since before the introduction of antimicrobials (D'Costa *et al.*, 2011). These genes are thought to have integrated into existing MGEs and transferred into pathogenic organisms, although various bottlenecks are thought to restrict this (Datta & Hughes, 1983; Martínez, 2011).

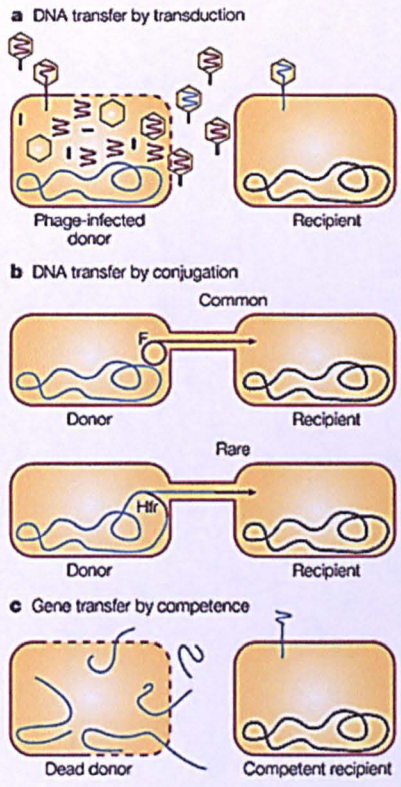
These environmental bacteria may act as reservoirs for resistance genes or mechanisms (Allen *et al.*, 2010; Cantón, 2009; Wright, 2010). Environmental bacteria are exposed to chemicals which may drive natural selection of such traits. In addition, resistance determinants have been shown to play a role in other cell functions of environmental bacteria (Martinez, 2009). Since the advent of the antimicrobial era, bacteria acquired resistance to newly introduced antimicrobials through HGT of MGEs such as plasmids (Ito *et al.*, 1999). Their natural selection appears in part to be driven by the widespread use and over-prescription of antimicrobials. Greater antimicrobial use enables those bacteria that harbour mutations or genes that confer antimicrobial resistance to survive and spread. However, the evolutionary processes driving resistance in environmental bacteria and hence, in part, in clinical pathogens are still largely unknown.

1.5. Genome of *S. aureus*

1.5.1. Mobile genetic elements (MGEs)

MGEs help drive the evolution of *S. aureus* by generating genetic variations between strains, thus enabling them to adapt to changing environments for survival. MGEs were first discovered in the 1940s (McClintock, 1950) and have since been found to be ubiquitous in both eukaryotes and prokaryotes. MGEs are segments of DNA that are capable of inter- and intra-cellular movement and may encode mechanisms of insertion into the bacterial chromosome or plasmid. Up to 20 % of the *S. aureus* genome is composed of MGEs which forms part of the accessory genome. MGEs also encode virulence and antimicrobial resistance factors, which may aid in the onset of disease, carriage or may cause complications in the treatment of infections (Holden *et al.*, 2008). The HGT of foreign DNA or genetic recombination in bacteria occurs via transformation, conjugation or transduction. HGT into *S. aureus* occurs predominantly from within the same species (Figure 7). This is partially due to bacteriophages being species-specific and due to mechanisms that minimise the transfer of DNA from other species (Holden *et al.*, 2008).

Figure 7. Methods of bacterial genetic transfer



a) Transduction is the incorporation of DNA from a bacterial genome into a bacteriophage, resulting in lysis of the donor cell. This incorporation may include random bacterial genes (generalised transduction) or specific bacterial genes (specialised transduction). This bacteriophage infects another recipient host and transfers this DNA fragment into the new host chromosome or plasmid. b) Conjugation involves the direct exchange of DNA between two cells via a pilus (Becker *et al.*, 2006). c) Transformation is the uptake of free DNA into the chromosome. This free DNA can be in the linear form however, the majority of successful transformations require circular DNA. Factors required for the competence of *S. aureus* however are as yet unknown (obtained from Redfield, 2001).

1.5.1.1. Bacteriophages

Bacteriophages are viruses and in S. *aureus*, most harbour a genome of approximately 45 kb. They can exist in the lytic state but most S. *aureus* bacteriophages exist in the lysogenic state where DNA is integrated into the chromosome and remains in a dormant state known as a prophage. They harbour short inverted repeats at the terminal ends and encode an integrase which specifies the insertion site and enables site-specific insertion into the chromosome (Carroll *et al.*, 1995). S. *aureus* genomes harbour up to four prophages which can either be common to all lineages or be lineage-specific. The spread of bacteriophages may be controlled by the host *Sau1* Restriction-Modification (R-M) system (see R-M system: section 1.6.3). In addition, some others such as the bacteriophage φ42 encodes its own R-M system such as the *Sau*42I R-M system (Dempsey *et al.*, 2005). Bacteriophages may convey a selective advantage as they can encode virulence factors and toxins such as PVL (Rahman *et al.*, 1992). Bacteriophages have been utilised for decades to type S. *aureus* strains based on their susceptibility pattern to a particular set of bacteriophages. They are also utilised to genetically manipulate S. *aureus in vivo*. Host DNA has been shown to integrate into prophages in the chromosome and prevent the formation of bacteriophage particles. This mechanism is thought to form SaPIs and may play a role in the co-evolution of bacteriophages and bacteria (Wirtz *et al.*, 2010; Yarwood *et al.*, 2002).

1.5.1.2. S. *aureus* pathogenicity islands (SaPIs)

SaPIs are approximately 15 kb and do not encode bacteriophage structural proteins, but use proteins encoded by bacteriophages present in the bacterium to excise from the

chromosome. SaPIs integrate into specific sites using an integrase and terminal inverted repeats under environmental conditions of stress (Lindsay *et al.*, 1998; Tallent *et al.*, 2007). *S. aureus* strains harbour up to two SaPIs which can be packaged into small phage head-like particles, released via cell lysis and can integrate into other cells (Tormo *et al.*, 2008). Like bacteriophages, SaPIs are thought to be species-specific and transfer is thought to be regulated by the *Sau1* R-M system. However, interspecies transfer has also been reported (Maiques *et al.*, 2007). As in bacteriophages, variations in SaPI integrases correlate with different insertion sites. Similarly, some SaPIs are relatively common and others are lineage-specific. SaPIs also display a mosaic structure, suggesting rearrangements with other SaPIs. SaPIs harbour R-M system proteins and virulence genes which can aid in the onset of disease (Baba *et al.*, 2008). Superantigens such as TSST-1, exotoxins and leukocidin components can be found on SaPIs, as can various enterotoxins and haemolysins (Baba *et al.*, 2008; Diep *et al.*, 2006; Fitzgerald *et al.*, 2001; Lindsay *et al.*, 1998).

1.5.1.3. Plasmids

Plasmids are circular sections of DNA up to 150 kb in size, which undergo autonomous replication. As *S. aureus* are less competent than other bacteria such as *Streptococcus pneumoniae*, plasmids are thought to be transferred by conjugation or transduction (Holden *et al.*, 2008). Strains of *S. aureus* harbour up to three plasmids which can either be free or integrated into the chromosome at sites such as the SCC*mec*. Plasmids from other species can be transferred into *S. aureus* and often carry antimicrobial resistance determinants and virulence factors. The plasmid pAM830 containing the *van*A transposon (Tn*1456*-like element) is thought to have transferred via conjugation from enterococci to form VRSA

(Flannagan *et al.*, 2003; Weigel *et al.*, 2003). High-level mupirocin resistance is associated with strains where the *mupA* gene, which encodes mupirocin resistance, is found on a plasmid (Ramsey *et al.*, 1996). Virulence factors such as enterotoxin B (Yamaguchi *et al.*, 2001) and exotoxins are also encoded on plasmids (Omoe *et al.*, 2003). Genes conferring resistance to bacteriocins have also been described on plasmids which may confer an advantage in mixed populations or under environmental stress (Navaratna *et al.*, 1999). In addition, point mutations in promoter regions can result in changes in gene expression that provide a selective advantage under conditions of selection due to antimicrobial therapy (Holden *et al.*, 2008; Nakaminami *et al.*, 2008).

1.5.1.4. Transposons

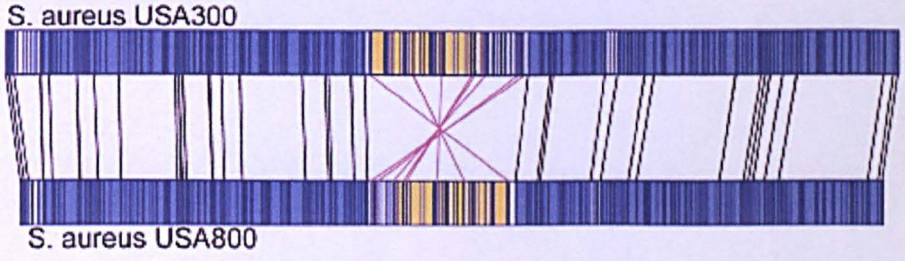
Transposons, sized up to 60 kb, are fragments of DNA that encode a transposase. This allows relocation of its related transposon to sites in the genome with which the transposon has no sequence homology. This can cause polar mutations (changes in gene expression downstream of the site of integration) and generate knockout mutations. Most *S. aureus* strains harbour one or two different transposons. Transposons harbouring antimicrobial resistance determinants for penicillin, erythromycin, tetracycline, aminoglycosides and vancomycin have been described in *S. aureus* (Holden *et al.*, 2008; Nakaminami *et al.*, 2008).

1.5.1.4.1. Insertion sequences (ISs)

ISs are a type of transposon which are relatively small in size (less than 2.5 kb) and only encode proteins required for transposition such as a transposase. They are found in most

bacterial species and over 1500 types are described to date in Eubacteria and Archaea (Siguier *et al.*, 2006). Most ISs harbour terminal inverted repeats. In *S. aureus*, ISs have been associated with individual lineages (Lindsay *et al.*, 2006). Multiple copies of an IS can be present throughout the chromosome and within plasmids. ISs flanking a piece of DNA can excise and transport it around the chromosome (composite transposon) (Mahillon & Chandler, 1998). This mechanism enables the exchange and recombination of DNA between plasmids and chromosomes. These may induce rapid rearrangement of the *S. aureus* chromosome and enable the bacterium to adapt to its environment. Integration of ISs in genes can disrupt cell functions, for example the integration of ISs into the lytic enzyme gene *lytH* results in increased meticillin resistance (Fujimura & Murakami, 2008). ISs have been utilised with other techniques to type *S. aureus* strains (Symms *et al.*, 1998).

Figure 8. Genetic inversion between USA300 and USA800 strains identified by whole-genome mapping (WGM)



The inverted region is highlighted in yellow and with purple lines between two optical maps (generated using whole-genome mapping, WGM) whilst the blue regions and black lines indicate homogeneous regions. White regions indicate those regions unique to each map (obtained from Shukla *et al.*, 2009).

Inversions in *S. aureus* are also rare and they are thought to occur when ISs flanking a region of DNA excise and reintegrate into the chromosome in the opposite orientation (Watanabe *et al.*, 2007). Only one large-scale inversion of approximately 500 kb has been described between *S. aureus* strains of USA300 and USA800 (Figure 8). This region was flanked by IS*1181* and harboured inverted repeats 73 bp in length (Shukla *et al.*, 2009). Inversions could alter the expression order and level of genes (Henderson *et al.*, 1999).

1.5.1.5. Genomic islands (GIs)

GIs are clusters of genes that can be transferred into the same *S. aureus* strain on multiple occasions from the same host (Roos & van Passel, 2011). Pathogenicity islands are those GIs that carry genes that encode functions that enhance virulence. All *S. aureus* strains harbour the pathogenicity islands, GIα and GIβ. These do not encode their own replication, excision and mobility and are thought to be transferred via transduction (Holden *et al.*, 2008). Virulence factors such as multiple toxin genes, leukocidins and bacteriocin biosynthesis genes have been described on these GIs (Gravat *et al.*, 1998; Jarraud *et al.*, 2001; Navaratna *et al.*, 1999). Although not all GIs encode virulence factors, differences in pathogenicity between *S. aureus* and *S. epidermidis* may be due to toxins encoded on these GIs (Gill *et al.*, 2005).

1.5.2. Staphylococcal cassette chromosome *mec* (SCC*mec*)

The *mecA* gene, which encodes PBP2A, is found on a large GI section known as the staphylococcal cassette chromosome (SCC) which is absent in MSSA (Beck *et al.*, 1986). This GI inserts in the chromosome at the *attB* site (bacterial chromosomal attachment site), the

sequence of which differs between MRSA and MSSA (Noto *et al.*, 2008). The *attB* site is located at the 3' end of an open reading frame which encodes an rRNA methyltransferase, *orfX*, located near the origin of replication of the chromosome; the *orfX* is conserved between strains (Boundy *et al.*, 2013; Ito *et al.*, 2001). SCC harbouring the *mecA* gene (SCC*mec*) and those that lack the *mecA* have been described. This GI can harbour other virulence factors and antimicrobial resistance determinants but lacks bacteriophage structural proteins and conjugative transfer (*tra*) genes. However, the mechanism of transfer of the SCC*mec* is still unknown (Hanssen & Sollid, 2006). Unless otherwise stated, the following information was obtained from the International Working Group on the classification of Staphylococcal Cassette Chromosome elements (IWG-SCC, 2009).

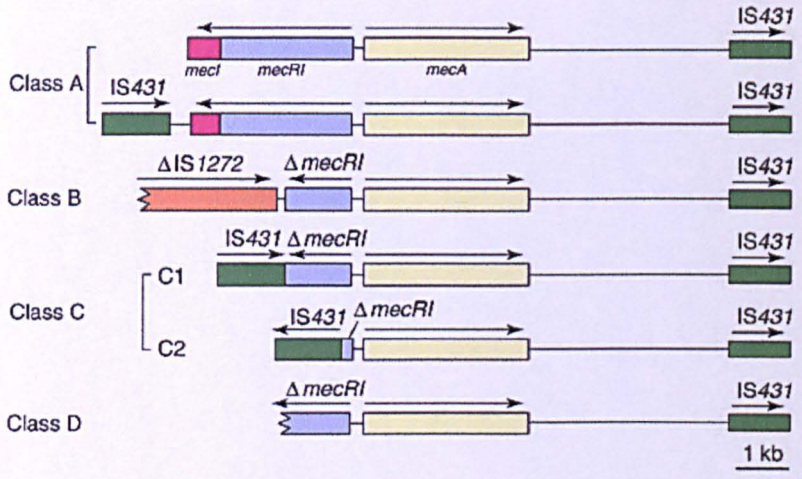
The SCC*mec* is composed of a *mec* complex, *ccr* complex and three joining regions (J regions) which surround the *mec* and *ccr* complexes. The left and right junctions of the SCC*mec* with the chromosome are known as the *att*L and *att*R respectively (Ito *et al.*, 1999). All SCC*mec* harbour direct repeats found next to the *att*L and within the SCC*mec* right terminal end. In addition, all SCC*mec* harbour inverted terminal repeats IR-L and IR-R (27 bp) found at the termini of the GI.

1.5.2.1. *mec* complex

The *mec* complex harbours the *mecA* gene, the associated regulatory genes (*mecI*, repressor and *mecR1*, transducer) and a variety of ISs (Ito *et al.*, 1999; Katayama *et al.*, 2000). The *mecI* and *mecR1* genes can be present in the functional or truncated form or can be absent. The *mec* gene complex is assigned to a class A to E based on the presence of genes and ISs

(Figure 9). *mec* complex class E is a recently described complex with the composition blaZ-*mecA*$_{LGA251}$-*mecR1*$_{LGA251}$-*mecI*$_{LGA251}$. The sequences of the *mec* and regulatory genes are divergent from those previously described and as a consequence, the *mec* gene has been designated *mecC* (García-Álvarez *et al.*, 2011).

Figure 9. Composition of *mec* complex: class A to D



The figure depicts *mecA*, its associated regulatory genes and insertion sequences which form the *mec* complex of the SCC*mec*. Class A composition: IS*431*-*mecA*-*mecR1*-*mecI*; class B composition: IS*431*-*mecA*-Δ*mecRI*-IS*1272*; class C1: IS*431*-*mecA*-Δ*mecRI*-IS*431* (both IS*431*s are arranged in the same direction); class C2: composition IS*431*-*mecA*-Δ*mecRI*-IS*431* (both IS*431*s are arranged in the opposite direction); class D: composition IS*431*-*mecA*-Δ*mecRI*. Δ, truncated sequence; IS, insertion sequence; *mecA*, meticillin resistant determinant; *mecI*, repressor; *mecR1*, transducer (obtained from Hiramatsu *et al.*, 2001).

1.5.2.2. *ccr* complex

The *ccr* gene complex encodes DNA recombinases (invertase-resolvase family) (Katayama *et al.*, 2000). Three recombinases have been described: *ccrA* and *ccrB* genes are found in the same *ccr* complex whilst *ccrC* is found in others. Both *ccrA* and *ccrB* are required for the site-specific integration and excision of SCC*mec* (Wang & Archer, 2010). The *ccrB* displays specificity for a 14 bp sequence (CGTATCATAAGTAA) at the terminal of *orfX* (Wang *et al.*, 2012). Based on the variations of these gene sequences, the genes are classified into one of eight allotypes (Table 1).

Table 1. Types of *ccr* complex

ccr complex	*ccr* genes
Type 1	*ccrA1* & *ccrB1*
Type 2	*ccrA2* & *ccrB2*
Type 3	*ccrA3* & *ccrB3*
Type 4	*ccrA4* & *ccrB4*
Type 5	*ccrC1*
Type 6	*ccrA5* & *ccrB3*
Type 7	*ccrA1* & *ccrB6*
Type 8	*ccrA1* & *ccrB3*

Allotypes of *ccrA* and *ccrB* or *ccrC* only which constitute *ccr* complex types 1 to 8 of the SCC*mec*.

1.5.2.3. Joining regions (J regions)

The J regions can harbour additional ISs and plasmid sequences which may carry antimicrobial resistance determinants. The majority of SCC*mec* types harbour three J

regions. The J1 region is between the right chromosomal junction and the *ccr* complex, J2 is the region between the *ccr* and *mec* complex and J3 is the region between the *mec* complex and the left chromosomal junction. J regions also harbour *pls*, a sequence which encodes a plasmin-sensitive surface protein, *dcs*, a downstream constant segment and *kdp* a sequence encoding for an adenosine triphosphate (ATP)-dependent potassium transport protein. The SCC*mec* site is a transposition hot spot where different plasmids and transposons have integrated or recombined (Diep *et al.*, 2006; Noto *et al.*, 2008). This has led to the emergence and dissemination of multiple antimicrobial resistant strains.

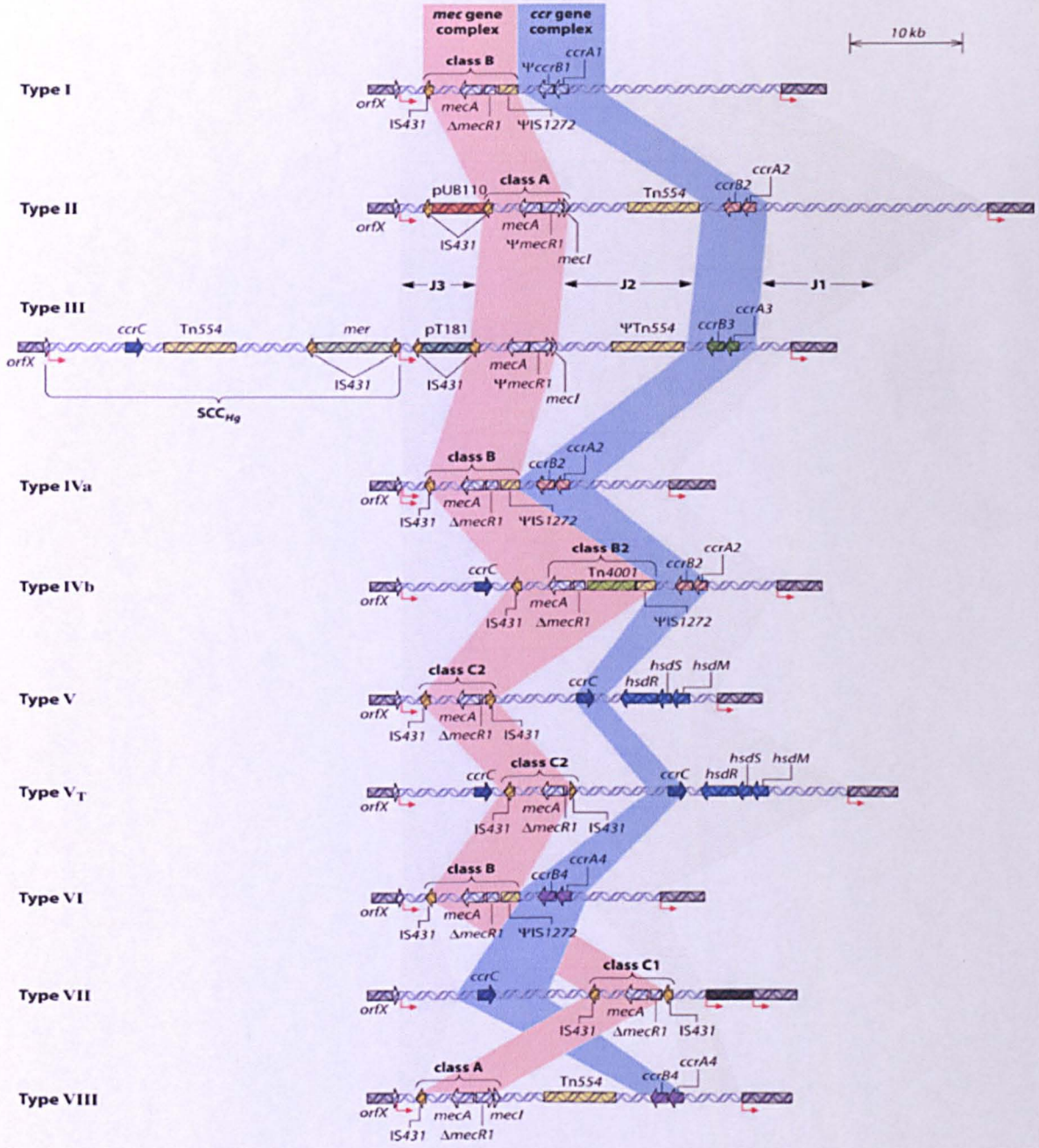
1.5.2.4. SCC*mec* types

The first SCC*mec* structure was described by Ito *et al.* (1999) and within two years, two more structures were identified (Ito *et al.*, 2001). By 2008, eight structurally diverse SCC*mec* types were described (Figure 10). Classification criteria for SCC*mec* types were established by the IWG-SCC in 2009. The classification of SCC*mec* is based on the class of *mec* complex and the type of *ccr* complex whilst variations in the J regions are used to define subtypes. The recombination of SCC*mec* types may occur and generate new SCC*mec* types (Zhang *et al.*, 2009). To date, 11 SCC*mec* types have been described (Table 2).

SCC*mec* types can be classified further into subtypes based on variations in their J regions. Differences in the J1 region are denoted as small letters after the SCC*mec* type e.g. IVa to IVh subtypes (Berglund *et al.*, 2009; Ito *et al.*, 2003; Ma *et al.*, 2002; Milheiriço *et al.*, 2007; Shore *et al.*, 2005). Variations due to the acquisition of MGEs are denoted by capital letters e.g. IIA and IVA whilst differences in the J1, J2 and J3 regions are denoted as numbers e.g.

II.1.1.1 and II.2.1.1. Composite SCC*mec* types that harbour two *ccr* complexes have also been reported (Heusser *et al.*, 2007; Higuchi *et al.*, 2008).

Figure 10. Staphylococcal Cassette Chromosome *mec* (SCC*mec*) types I to VIII



Schematic diagram depicting *mec* and *ccr* complex and features identified in the J1, J2 and J3 regions which constitute SCC*mec* types I to VIII (obtained fromDavid & Daum, 2010).

Table 2. Types of Staphylococcal Cassette Chromosome *mec* (SCC*mec*)

SCC*mec* type	*ccr* complex	*mec* complex	Reference*
I	1	B	(Ito *et al.*, 2001)
II	2	A	(Ito *et al.*, 1999)
III	3	A	(Ito *et al.*, 2001)
IV	2	B	(Ma *et al.*, 2002)
V	5	C2	(Ito *et al.*, 2004)
VI	4	B	(Oliveira *et al.*, 2006)
VII	5	C1	(Berglund *et al.*, 2008)
VIII	4	A	(Zhang *et al.*, 2009)
IX	1	C2	(Li *et al.*, 2011)
X	7	C1	(Li *et al.*, 2011)
XI	8	E	(García-Álvarez *et al.*, 2011)

Combination of *mec* and *ccr* complex found in SCC*mec* types I to XI.

***References indicate the first study to describe each type.**

1.5.2.5. SCCnon-*mec*

Different SCC elements that lack the *mecA* gene but harbour other resistance or virulence factors have been described. The SCC*capl* lacks *mecA* and harbours the *capl* operon which confers resistance to phagocytosis, indicating this element may confer a selective advantage (Luong *et al.*, 2002). Further elements such as the SCC*fur* and SCC*Hg* (Figure 10) have been described which confer resistance to fusidic acid and mercury.

1.5.3. Determining the sequence of the *S. aureus* genome

Genome sequences enable greater understanding of the biological processes of the cell. They enable us to shed new light on bacterial growth, replication, survival and adaption as well as aiding in the development of new typing techniques. Sanger and his colleagues proposed the di-deoxynucleotide (ddNTP) sequence determination method in 1975, which is now utilised

globally (Sanger & Coulson, 1975). This method involves amplification of target DNA using a mixture of deoxynucleotides (dNTP) and ddNTPs. The latter of which terminates chain elongation. This reaction is repeated four times, once for each ddNTP (A, T, C or G). Each reaction is separated by gel electrophoresis in individual wells and the size of the fragments determined by the distance travelled in the gel. The recent developments of next-generation sequencing (NGS) technologies are based on sequencing-by-synthesis method, which reduces the speed and cost and could lead to NGS being utilised in routine laboratories in the future.

As a major human pathogen, *S. aureus* has been a focus for NGS studies. Over 180 fully annotated genomes have been sequenced to date and are available on the NCBI database (http://www.ncbi.nlm.nih.gov). The *S. aureus* genome is composed typically of a 2.7-3.0 Mb chromosome and one or more plasmids and has a G+C content between 30 and 39 % (Dale & Park, 2004; Takeuchi *et al.*, 2005). Approximately 82-90 % of the chromosome is coding DNA, forming 2500-2900 protein coding sequences (Baba *et al.*, 2008; Lindsay *et al.*, 2008).

The sequenced strains belong to approximately 15 clonal complexes (CCs) which are based on MLST (see MLST: section 1.8). Comparative studies have revealed approximately 75 % of the genome is conserved between strains. These regions encode essential cell functions such as growth and metabolism. A large proportion of the variable part of the genome is composed of transposons, ISs, SaPIs and bacteriophages which play a role in antimicrobial resistance, virulence and pathogenicity (Table 3) (Fitzgerald *et al.*, 2001).

Table 3. Basic features of whole-genome sequences of *S. aureus*

	Strain designation											
	N315	Mu50	MW2	MRSA252	MSSA476	COL	USA300	Newman	NCTC 8325	RF122	JH1	JH9
Length of sequence (bp)	2814816	2878040	2820462	2902619	2799802	2809422	2872769	2878897	2821361	2742531	2906507	2906700
G + C content	32.8 %	32.9 %	32.8 %	32.8 %	32.9 %	32.8 %	32.8 %	32.9 %	32.9 %	32.8 %	33.0 %	33.0 %
Protein coding regions	2593	2714	2632	2671	2565	2673	2560	2614	2892	2589	2747	2697
IS*	20	23	6	30	5	10	12	12	8	11	17	17
Transposon Tn*554*	5	2	0	3	0	0	0	0	0	0	2	2
Bacteriophage	1	2	2	2	2	1	2	4	3	1	4	4
SaPI†	3	4	4	1	0	5	6	4	4	5	3	3
Plasmid	1	1	1	0	1	-	3	-	-	-	-	-
Reference	(Baba *et al.*, 2002; Kuroda *et al.*, 2001)			(Holden *et al.*, 2004)		(Baba *et al.*, 2008)						

This table displays size and G+C content of the genome, number of protein coding regions and mobile genetic elements.

***IS, insertion sequence.**

†SaPI, *S. aureus* pathogenicity island.

1.6. Evolution of *S. aureus*

1.6 1. Processes of evolution

Bacteria evolve via the action of natural selection upon genetic variation. Genetic variations can arise via random point mutations or via the HGT of foreign DNA. Over time, we expect mutations that are advantageous to increase in frequency via natural selection, thus enabling the bacteria to adapt to their environment. Point mutations can result in changes in virulence and antimicrobial resistance. Studies have shown, for example, that mutations in the isoleucyl-transfer RNA synthetase and *fusA* gene result in increased mupirocin and fusidic acid resistance respectively (Antonio *et al.*, 2002; Besier *et al.*, 2003). In addition, mutations upstream of *parE* result in a change in gene expression of topoisomerase IV which is associated with quinolone resistance (Ince & Hooper, 2003). Natural selection may maintain the sequence of genes encoding essential cell functions such as those utilised in MLST (see MLST: section 1.8).

1.6.2. Restriction-Modification (R-M) systems

The presence of genes encoding surface-associated proteins and regulators of virulence genes have been shown to vary between lineages but are relatively stable within a lineage (Lindsay *et al.*, 2006). This suggests mechanisms within *S. aureus* enable the HGT of DNA within a lineage at a higher frequency than between lineages. The hypothesis that HGT is restricted between lineages supports the evolutionary model for the emergence of MRSA. The insertion of the SCC*mec* transforms MSSA strains into MRSA (Ito *et al.*, 1999). There have been limited insertions of the SCC*mec* into MSSA lineages (Robinson & Enright, 2003).

MRSA therefore display a highly clonal population structure and the majority of isolates therefore belong to a limited number of closely related genotypes or CCs.

One such mechanism, R-M systems, has been highlighted as an important factor probably influencing its evolution. R-M systems are thought to control the uptake of foreign DNA from other *S. aureus* lineages or species by minimising the uptake of DNA such as that of bacteriophage. R-M enzymes recognise host-specific sequences and carry out modifications such as methylation; the unmodified foreign sequences are then digested by a restriction endonuclease. Restriction endonucleases cleave the DNA at a specific sequence usually between 4 to 8 bp long. Four types of R-M systems have been identified in bacteria (Waldron & Lindsay, 2006). Type I *sau1* R-M system has been identified in all sequenced strains of *S. aureus*. This system is composed of three types of genes: one copy of a restriction enzyme subunit *sau1hsdR*, two copies each of a modification enzyme subunit *sau1hsdM1* and *sau1hsdM2* and two copies each of a specificity enzyme subunit *sau1hsdS1* and *sau1hsdS2*. The genes encoding for each of the corresponding modification and sequence specificity subunits are found on operons on GIα and GIβ.

A study by Waldron and Lindsay (2006) identified a stop codon in the *sau1hsdR* gene that enabled the uptake of *E. coli* plasmids. The *sau1hsdR* and both *sau1hsdM* genes are highly conserved between different strains. However, two highly variable DNA binding regions named TRD1 and TRD2 were identified within the conserved regions of the *sau1hsdS*; five and nine subtypes of these regions were identified respectively. Different combinations of each of the two subtypes correlated with an *S. aureus* lineage (CC8, 12, 15, 22, 25, 39, 45 or 51). This indicates isolates of different lineages recognise and modify a different sequence which may explain why HGT occurs at a higher frequency within lineages than between

lineages. Variation in these genes may therefore be involved in the establishment of new lineages (Waldron & Lindsay, 2006). A study by Veiga and Pinho (2009) showed deletion of the *sau1hsdR* gene resulted in no increase in the ability of *S. aureus* to take up plasmids or bacteriophages. This implies additional factors such as a second R-M system contributing to the acceptance of foreign DNA into *S. aureus*. Other R-M systems such as the *Sau*421 R-M system have been described in *S. aureus*. They display restriction endonuclease activity but further work is needed to identify their specific recognition sequence (Dempsey *et al.*, 2005).

1.6.3. Evolution of MRSA lineages

1.6.3.1. Origin of SCC*mec*

The origin of SCC*mec* is unknown. Archer *et al.* (1996) proposed the integration of *mecA* into the ancestral SCC*mec* occurred from *Enterococcus hirae* via *S. haemolyticus* and *S. epidermidis*. A *mecA* homologue has been described in *S. sciuri*, which is thought to be the most ancient *Staphylococcus* species. However, isolates with this gene show little meticillin resistance (Couto *et al.*, 2000). A proportion of *S. sciuri* isolates (46 %) harbour a *mecA* gene that encodes PBP2A identical to that found in MRSA and confers meticillin resistance. It is thought to be the ancestor of the *mecA* in MRSA (Couto *et al.*, 2000). Before the introduction of meticillin, the ancestral form of SCC*mec* lacked *mecA* (Fluit & Schmitz, 2003). Shore *et al.* (2008) supports this theory and states that the function of such elements may have been the transfer of DNA between staphylococci under conditions of selection due to environmental stress, apart from antimicrobial pressure (Katayama *et al.*, 2003a). Since the acquisition of *mecA* however, the transfer of the SCC*mec* may be driven by meticillin selective pressure. Although the exact origin of SCC*mec* is unknown, there is evidence that

CoNS such as *S. epidermidis* act as a reservoir and can transfer the SCC*mec* to *S. aureus* (Hanssen & Sollid, 2006; Tulinski *et al.*, 2012).

1.6.3.2. Insertion of SCC*mec*

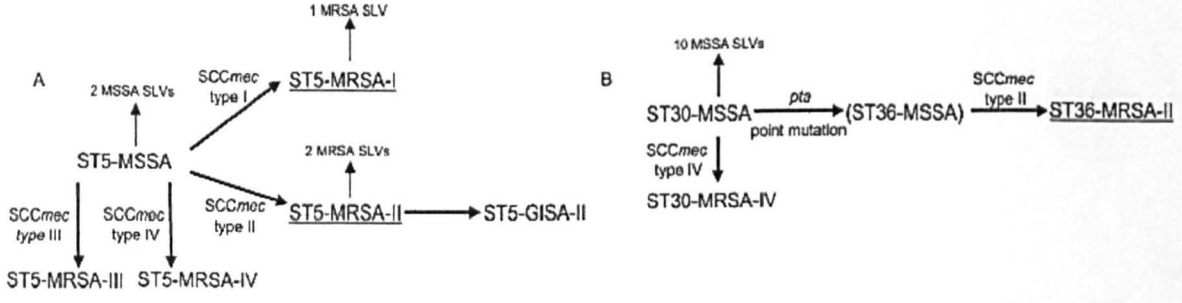
Two theories postulate the insertion of the SCC*mec* into MSSA to form MRSA. One theory is a single insertion event whereby all MRSA have evolved from a single origin or clone (a collection of organisms derived asexually from a single individual and with a common genotype) (Kreiswirth *et al.*, 1993; Lacey & Grinsted, 1973; van Belkum *et al.*, 2007). Based on the genetic diversity of MRSA exhibited by multilocus enzyme electrophoresis, Musser and Kapur (1992) proposed the second theory that multiple independent insertions of the SCC*mec* element have occurred. Whole-genome variation in MRSA has also been investigated by DNA microarrays which cover approximately 90 % of the genome (Fitzgerald *et al.*, 2001). This revealed evolution of five genotypically divergent groups that did not share a common ancestor, indicating independent acquisitions of the SCC*mec* into MSSA. This has led to the wide-spread acceptance of this multi-clone theory.

1.6.3.3. Evolutionary models for the origin of major MRSA lineages

MLST has revealed that human MRSA belong to one of 11 major clones or 5 CCs, based on their sequence type (ST) and SCC*mec* type (Enright *et al.*, 2002). Enright *et al.* (2002) observed that with the exception of ST5 and ST8, the majority of STs were associated with one SCC*mec* type. They were able to propose evolutionary models for the five major MRSA CCs associated with HA infections: CC5, 8, 22, 30 and 45. MSSA ST5, 8, 30 and 45 isolates were present in the 1960s prior to the emergence of EMRSA clones, indicating these

lineages were successful colonisers and causes of invasive disease prior to their acquisition of the SCC*mec* (Oliveira *et al.*, 2001). In contrast, community-associated MRSA (CA-MRSA) belong to distinct combinations of genetic background and SCC*mec* type. The origin of CA-MRSA has been postulated to be the transfer of *mecA* from a HA-MRSA strain into a susceptible strain of a divergent genetic background (Chambers, 2001). CA-MRSA with PVL has been described in six main genetic backgrounds: ST1, 8, 30, 59, 80 and 93.

Figure 11. Evolutionary model for the origin of CC5 and CC30

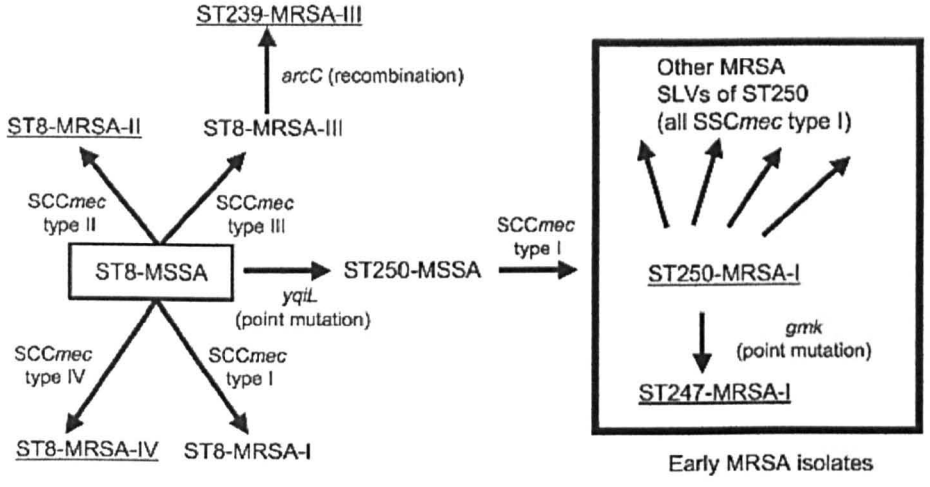


(A) Proposed evolutionary origin of major epidemic MRSA (EMRSA) clones of ST5. SCC*mec* types I to IV have inserted into ST5 and the majority of glycopeptide-intermediate *S. aureus* (GISA) isolates have emerged from the ST5-MRSA-II clone. (B) Proposed evolutionary origin of major EMRSA clone of ST36. ST36-MRSA-II (EMRSA-16) evolved from an ST36-MSSA progenitor within which the SCC*mec* type II has inserted. ST36 is a single-locus-variant (SLV) (point mutation in *pta* locus) of ST30-MSSA. The SCC*mec* type IV has also been acquired once into ST30 (obtained from Enright *et al.*, 2002).

All five major hospital-associated MRSA (HA-MRSA) clones were derived from the insertion of the SCC*mec* into their MSSA progenitors. Multiple insertions of the SCC*mec* have been proposed within CC5 and 30 (Figure 11). A single insertion event of SCC*mec* type IV into

ST22 (EMRSA-15) has led to the emergence of a successful MRSA clone. The term successful clone refers to a clone that has given rise to epidemics in the past. Conversely, SCC*mec* type IV and SCC*mec* type II has been acquired by ST45 on four and one occasion respectively. The pandemic 80/81 clone epidemic in the 1950s evolved from ST30-MSSA and subsequently acquired the PVL toxin (Feil & Enright, 2004).

Figure 12. Evolutionary model for the origin of CC8



ST250 with SCC*mec* type I evolved from an ST8-MSSA progenitor within which a single point mutation in the *yqiL* locus gave rise to ST250-MSSA. Independent insertions of SCC*mec* type I, II, III and IV have occurred within ST8-MSSA of which ST8-MRSA-II and IV have subsequently given rise to successful MRSA clones. ST8-MRSA-III rather than ST8-MSSA was proposed to be the ancestor of ST239-MRSA-III (previously Brazilian clone) as all ST239 isolates were MRSA (obtained from Enright *et al.*, 2002).

CC8 was found to include the original MRSA clone (Archaic clone) ST250 with SCC*mec* type I (Enright *et al.*, 2002). The evolutionary model of CC8 was proposed by Robinson and Enright (2003) (Figure 12). ST250-MSSA isolates were reported in Denmark from the 1960s (Crisóstomo *et al.*, 2001). ST247-MRSA-I (previously Iberian clone) was proposed to have

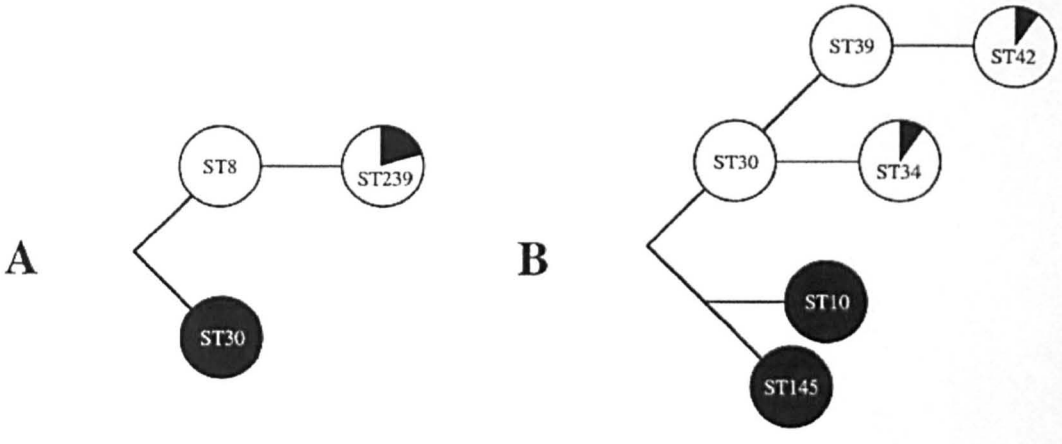
evolved from ST250-MRSA-I via a point mutation at the *gmk* locus. ST257-MRSA-I, another early MRSA clone, also arose from ST250-MRSA-I via a point mutation at the *glpF* locus (Robinson & Enright, 2003), whilst ST254-MRSA-I (EMRSA-10) is thought to have arisen directly from ST8-MSSA.

Certain CCs such as ST25 and ST121 are composed predominantly of MSSA while others are predominated by strain types that include important MRSA strains such as ST36 (notable for including EMRSA-16 isolates) (Enright *et al.*, 2002). This implies that either certain lineages such as ST121 have a genetic background within which acquisition of the SCC*mec* is not stable or the combination of genetic lineage and SCC*mec* element is not successful. However, these CCs may encode factors associated with growth or survival which make it more transmissible or persistent (Gomes *et al.*, 2006). The SCC*mec* type present in strains was specific to the genetic background of the clone (Oliveira *et al.*, 2001). SCC*mec* type I and III were found in strains from the same genetic background and SCC*mec* type II was found in strains from a different genetic background. However, SCC*mec* type IV was found in strains from both backgrounds. The latter type may be acquired frequently as its small size makes the transfer more efficient or it may have a lower fitness cost on the strain, making its transfer more favourable (Ma *et al.*, 2002). Once SCC*mec* is acquired within an MSSA ST, it is retained and is rarely replaced by another SCC*mec* type (Robinson & Enright, 2003). Studies have suggested that SCC*mec* was only stable in certain genetic backgrounds and was limited in its acquisition which could explain the clonal nature of MRSA (Katayama *et al.*, 2003b; 2005). Further research in this area is needed to provide an insight on the regulation of other MGE insertions and the limited spread of vancomycin resistance.

1.6.3.4. Large-scale insertions

Large-scale insertions into *S. aureus* genomes are relatively rare. Based on the sequences of seven MLST loci and surface proteins, a large-scale recombination event between the progenitors of ST8 and ST30 has been predicted to form ST239, which displays a longer cell cycle (Figure 13) (Robinson & Enright, 2003). Of 41 surface protein genes, 22 genes (approximately 557 kb) were similar to ST30 whilst 19 genes (approximately 2220 kb) were similar to ST8.

Figure 13. Evolutionary relationships of sequence types involved in large chromosomal replacements



The black and white sections represent the approximate chromosome contributions from each parent to form the mosaic chromosome. A) ST239 has descended from ST8 and ST30 progenitors. B) ST34 has descended from ST30 and ST10 and ST145 progenitors. ST42 has descended from ST39 and ST10 and ST145 progenitors (obtained from Robinson & Enright, 2004).

A second large-scale rearrangement of ST30 and ST39 with ST10 and ST145 has also been described (Figure 13). Approximately 244 kb and 256 kb from ST10 and ST145 have been exchanged into ST30 and ST39 to form ST34 and ST42 respectively. As these rearrangements have only been observed in CC30, this implies lineage-specific genes may be essential (Robinson & Enright, 2004). In addition, due to the large nature of the DNA exchange, this transfer is likely to occur by conjugation.

1.6.4. Hospital and Community-associated MRSA

The first outbreak of MRSA was reported in 1963. However, it was not until the 1980s that the pandemic of MRSA truly began. The spread of MRSA mirrors that of penicillin-resistant *S. aureus* decades previously, with cases initially emerging in hospitals and later in the community (Chambers, 2001). The incidence of MRSA infections vary worldwide. MRSA infections are however associated with up to 50 % higher mortality rate as compared to MSSA infections (Cosgrove *et al.*, 2003; Hanberger *et al.*, 2011; Yoon *et al.*, 2005). This may be due to the limited antimicrobial therapy available for MRSA, rather than increased virulence (Lowy, 2003).

MRSA infections have historically been HA. However, in the last two decades MRSA-associated infections in the community have risen dramatically. These CA-MRSA strains are not associated with the same risk factors as HA-MRSA strains and often affect all individuals, irrespective of age or health (Zetola *et al.*, 2005). The risk factors associated with CA-MRSA strains are intravenous drug use (Qi *et al.*, 2005), men who have sex with other men (Sztramko *et al.*, 2007) and members of contact sport teams (Centers for Disease Control and Prevention, 2003). HA-MRSA infections are defined as those that present symptoms 48 h after admission to the hospital, whereas CA-MRSA infections are those that

present symptoms within 48 h and are acquired from settings unrelated to hospitals such as schools and gyms (Buck *et al.*, 2005). This definition is widely accepted despite the fact that CA-MRSA strains have been noted to show symptoms >48 h after hospital admission, particularly amongst patients with a history of exposure to the hospital (Wilson *et al.*, 2011).

CA-MRSA strains differ from HA-MRSA in the diseases they cause, their antimicrobial susceptibilities as well as their genetic background (Cooke & Brown, 2010). Respiratory tract, urinary tract and wound infections tend to be due to HA-MRSA whilst CA-MRSA infections often present as severe skin and soft tissue infections, necrotising pneumonia and bacteraemia. HA-MRSA strains are often resistant to β-lactams and multiple classes of antimicrobials whilst CA-MRSA strains tend to be susceptible to most classes of antimicrobials excluding the β-lactams. This may be attributed to lower antimicrobial selective pressure in the community compared to the healthcare environment. A higher proportion of community strains are also PVL-positive and belong to distinct genetic backgrounds (Chambers, 2001). SCC*mec* types IV and V tend to be attributed to community strains whilst types I, II and III are traditionally associated with HA-MRSA. Whether community strains are more virulent is debatable; however studies have suggested patients with infections as a result of community strains are more likely to develop severe sepsis (Kreisel *et al.*, 2011). Colonisation by CA-MRSA strains can replace the existing colonising MSSA strains. MSSA however still account for the majority of *S. aureus* infections and are still a major clinical concern (Mostofsky *et al.*, 2011).

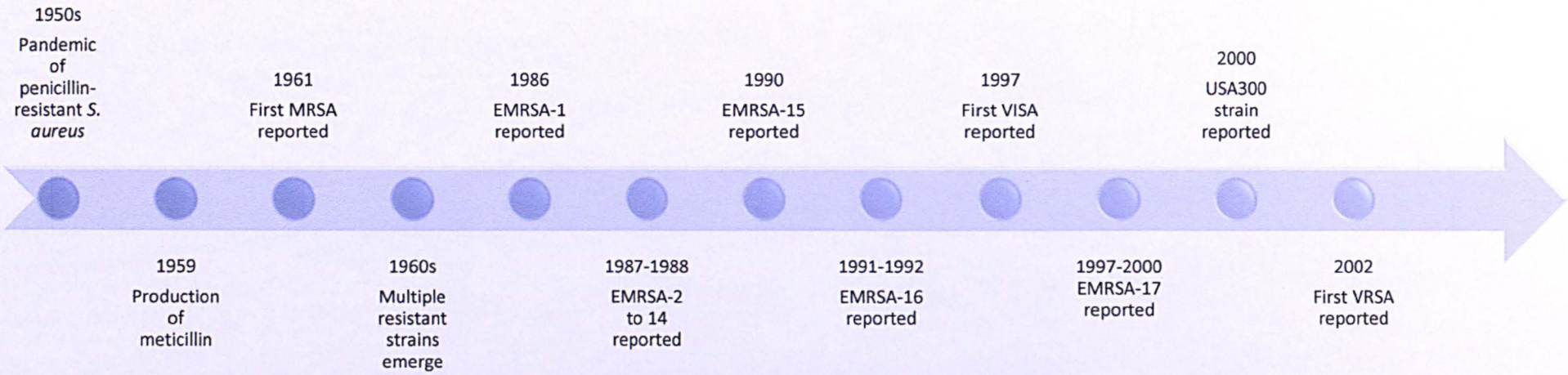
1.6.5. Waves of epidemic *S. aureus*

After the emergence of MRSA in 1961, strains resistant to multiple antimicrobials were reported throughout the 1960s (Fluit & Schmitz, 2003). This coincided with the decline of

penicillin-resistant strains such as bacteriophage type 80/81 that were prevalent in the 1950s. In the 1970s, resistance to aminoglycosides such as gentamicin was also reported. MRSA re-emerged simultaneously as strains that express multiple resistance phenotypes in the USA during the 1980s. In the UK, the first EMRSA strain (EMRSA-1), defined by the ability to spread rapidly amongst patients (Hospital Infection Society and British Society for Antimicrobial Chemotherapy, 1990), was isolated in 1986 (Marples *et al.*, 1986). Thirteen further epidemic strains, EMRSA-2 to EMRSA-14, were defined by Staphylococcal Reference Laboratory (SRL) at Public Health England, PHE, London, (previously known as Public Health Laboratory Service and Health Protection Agency) based on strains received over a six month period from 1987 to 1988 (Kerr *et al.*, 1990).

EMRSA-15 was first isolated in the Midlands and South East of England and EMRSA-16 was detected a year later, between 1991 and 1992, in outbreaks at three Northampton hospitals (Cox *et al.*, 1995; Richardson & Reith, 1993). EMRSA-15 and EMRSA-16 are the predominant strains in the UK. This is in accordance with studies that suggest one or two MRSA lineages dominate in one geographical area (Figure 14) (Cockfield *et al.*, 2007). These two strains have subsequently spread widely in the UK and have been reported in other European countries (Cookson, 2011; Gould *et al.*, 2008), Australia (Pearman *et al.*, 2001), Asia (Hsu *et al.*, 2007) and USA (McDougal *et al.*, 2003). EMRSA-15 and EMRSA-16 are commonly resistant to erythromycin, ciprofloxacin and to β-lactams (Ellington *et al.*, 2010).

Figure 14. Emergence and waves of epidemics of antimicrobial resistant *S. aureus*



Hospital-associated MRSA emerged in 1961 since then, numerous epidemic MRSA (EMRSA) strains have been a major clinical concern in the UK. Community-associated MRSA, vancomycin-intermediate *S. aureus* (VISA) and vancomycin-resistant *S. aureus* (VRSA) have emerged in the past few decades and continues to increase in prevalence.

By the late 1980s and 1990s, a number of pandemic clones defined by pulsed-field gel electrophoresis (PFGE) were circulating around the world. These included the Brazilian, Hungarian, Iberian, New York/Japan and Pediatric clones. The Iberian clone is closely related to the Archaic clone, thought to be the earliest MRSA clone in Europe, which is now extinct (de Lencastre *et al.*, 2000). At the beginning of the 21st century, EMRSA-17 emerged as a multi-resistant strain isolated predominantly from the south coast of the UK and London (Aucken *et al.*, 2002). This period also saw the emergence of PVL-positive strains such as USA300 which has spread rapidly amongst the community and in hospital settings in North America (Reyes *et al.*, 2009). Recent cause for concern has been the transmission of strains into humans from livestock which are thought to be a reservoir for MRSA. Cases of skin and soft tissue as well as joint infections have been reported in people with close contact to pigs. Isolated infections in patients with no identifiable link to livestock have also been reported (Krziwanek *et al.*, 2009; Mammina *et al.*, 2010). Despite the emergence of these clones, a limited number of MRSA clones have spread in Europe (Grundmann *et al.*, 2010).

1.7. Bacterial identification and typing

1.7.1. Concepts of bacterial typing

Bacterial typing techniques use phenotypic or genotypic characteristics to discriminate between isolates of the same species. Typing is useful in bacterial epidemiology, infection control measures and in the surveillance of infectious diseases on a local and global scale. Surveillance can warn us of emerging strains and clones that may cause future outbreaks. On a global scale, typing isolates from different origins may enable the population structure of the species to be determined and the significance of the diversity understood. Genetic typing

methods are therefore commonly utilised in the identification of markers and mechanisms related to virulence, which could be utilised in diagnostics or as future drug targets (Fluit & Schmitz, 2003; Trindade *et al.*, 2003; van Belkum *et al.*, 2007).

In this context, an isolate is defined as a population of bacterial cells in pure culture derived from a single colony. A strain is defined as descendants of a single isolation from a pure culture which is distinguished from other bacteria of the same species by their phenotypic or genotypic characteristics. A clone is defined as a collection of organisms derived asexually from a single individual and with a common genotype (van Belkum *et al.*, 2007). However, individual members of a clone may vary through spontaneous mutations. Prompt detection and identification of bacteria from clinical samples often requires a pure culture. The effectiveness of a typing technique is evaluated and validated against two types of criteria: performance criteria, which survey the efficacy of the technique and convenience criteria, which explore the efficiency. Performance criteria include stability of the marker, typeability, discriminatory power which is the ability of the technique to distinguish between two unrelated isolates and reproducibility, which is the ability to assign an isolate to the same type on independent occasions. Convenience criteria include the rapidity and flexibility of the technique, accessibility of the technique to laboratories, cost and the ease of the interpretation, comparison and the geographical and temporal portability of data (Trindade *et al.*, 2003; van Belkum *et al.*, 2007).

1.7.2. Identification of *S. aureus*

The differentiation between microorganisms from clinical samples includes determination of phenotypic and genotypic characteristics and recently the characteristics of proteins in

bacteria such as antimicrobial susceptibility testing through to 16S rRNA gene sequence determination and mass spectrometry.

1.7.2.1. Antimicrobial susceptibility testing

Antimicrobial susceptibility testing can have direct uses in clinical therapy. Isolates are tested against a range of antimicrobials forming an antimicrobial resistance pattern or antibiogram. Lineages of *S. aureus* can exhibit susceptibilities to different antimicrobials (Holden *et al.*, 2008). The stability and discriminatory power of this technique is variable, in part due to the resistance determinants being encoded on MGEs. Thus the discriminatory power of the susceptibility profiling for typing purposes is limited (van Belkum *et al.*, 2007).

1.7.2.2. Bacteriophage typing

Bacteriophages, which were discovered in 1915 (Twort, 1915), have been utilised for the typing of bacteria . Bacteriophages are restricted to the organisms they are able to infect and can differ between the genus and strain level (Dale & Park, 2004). Typing is based on the susceptibility patterns generated when cultures are exposed to a diverse set of bacteriophages (Holden *et al.*, 2008). Bacteriophage typing specific for staphylococci was proposed in the early 1950s and since then, this technique has become popular in reference centres due to the low cost and ease of use. However, this does not account for the cost of the curation and maintenance of the bacteriophage collection (Blair & Williams, 1961; Singleton, 2004). The technique however has low resolution, displays poor reproducibility and cannot distinguish

between strains. In addition, the quality control of bacteriophages is time-consuming and laborious (van Belkum *et al.*, 2007).

1.7.2.3. Multilocus enzyme electrophoresis (MLEE)

Multilocus enzyme electrophoresis (MLEE) utilises differences in the electrophoretic mobility of 15 to 25 housekeeping enzymes. The mobility varies, depending on the protein size and charge which are caused by polymorphisms in the genes encoding the housekeeping proteins. The mobility of the enzyme is directly related to the allele at the locus that encodes the enzyme (Selander *et al.*, 1986). The reproducibility of the technique is variable as different enzymes can have the same electrophoretic mobility. MLEE has been particularly useful for studies investigating the population structure of MRSA and have revealed multiple insertions of the SCC*mec* into *S. aureus* (Musser & Kapur, 1992). However, some polymorphisms do not cause changes in the amino acid sequence, especially in housekeeping genes, where the majority of mutations are neutral. In addition, post-translational modifications of enzymes may bias the level of variation estimated (Selander *et al.*, 1986).

1.7.2.4. Determining the sequence of the 16S ribosomal RNA (rRNA) gene

Determining the sequence of the 16S rRNA gene is a powerful identification technique for most bacteria and was first discovered by Carl Woese (Woese *et al.*, 1975). The gene is composed of highly conserved regions alternating with hypervariable regions. Sequence determination of the hypervariable regions is possible by designing primers complementary to the highly conserved flanking regions (Dale & Park, 2004). The rRNA genes are ideal as

genetic markers because they are ubiquitous and their function is conserved over time and can also be used as a measure of evolution. The number of differences can therefore infer evolutionary distance and relationships between species via the construction of phylogenetic trees. The 16S rRNA gene has been used for bacterial identification but there is no universal agreement on what percentage similarity constitutes a species. Almost all isolates can be assigned to the genus level and up to 83 % to the species level (Janda & Abbott, 2007). Problems arise due to similar sequences between species, poor sequences deposited in nucleotide databases and novel sequences (Ashelford *et al.*, 2005).

1.7.2.5. Mass spectrometry

The genome of an organism is considered to be stable. In contrast, the proteome changes in response to the organism's environment and proteomics identifies proteins expressed under specific growth conditions or at a particular growth stage (Holden *et al.*, 2008). Mass spectrometry is used to determine the molecular composition of a sample or mixture. The principle of time-of-fight mass spectrometer was first described by William Stephens in 1946. However, it was two years later when the first spectrometer was designed and spectra were published (Cameron & Eggers, 1948; Stephens, 1946). Matrix-assisted laser desorption ionization-time-of-flight mass spectrometry (MALDI-TOF MS; will be referred to as MALDI here on) is a technique that enables the analysis of biomolecules such as DNA and proteins. The basic principle of MALDI involves overlaying the sample (target spot) with a matrix solution and subjecting it to an intense laser. The matrix solution initially absorbs this energy and is ionised. This aids in the ionisation of the sample by acting as a proton acceptor or donator to or from the molecules. The vaporised sample (formed of ions) are directed into a flight tube which is under vacuum and is subjected to an electromagnetic field. This

accelerates and deflects the ions, based on their mass-to-charge ratio. Smaller ions are distorted to a greater extent than larger ions and reach the detector (at the end of the flight tube) first. A signal is generated when an ion reaches the detector giving rise to a spectrum.

MALDI is used routinely in hospital laboratories as it produces spectra that provide a molecular fingerprint for the detection of bacterial species. MALDI can identify the staphylococci down to the species level as it is based on the identification of ribosomal peptides (Carbonnelle *et al.*, 2007). A more recent study has utilised MALDI to discriminate between MRSA belonging to five major CCs (CC5, 8, 22, 30 and 45) associated with HA-MRSA infections (Wolters *et al.*, 2011). In addition, MALDI has been utilised to identify isolates producing the PVL toxin and delta-toxin (Bittar *et al.*, 2009; Gagnaire *et al.*, 2012). In both these cases, the assignment is based on the detection of a specific mass-to-charge peak. This technique is rapid and reproducible and can be utilised for the rapid assignment of MRSA lineages. Current topics of research include using MALDI to identify antimicrobial resistance (Wieser *et al.*, 2012).

1.8. Genotyping of *S. aureus*

Advances in molecular biology have led to the development of DNA-based typing techniques which are rapid and generally reproducible. These techniques can provide additional information about the population structure and genetic diversity between isolates. These techniques are less prone to changes due to environmental influences and genotypic markers represent stable targets that enable the application of molecular techniques to all strains.

1.8.1. Plasmid profiling

Plasmid profiling was the first molecular technique used for MRSA typing (Trindade *et al.*, 2003). This technique involves extraction of plasmid DNA, which may or may not be digested with restriction endonucleases and separation on an agarose gel to obtain a profile. However, as plasmids can be lost and acquired, related isolates can display different profiles. In addition, plasmids can harbour transposons or ISs which can be excised and enter the chromosome, thus altering the plasmid (Hartstein *et al.*, 1995). It is therefore difficult to ascertain whether dissemination of the clone or the plasmid is occurring (Tenover *et al.*, 1994). The reproducibility and discriminatory power are variable due to the limited number of plasmids, the stability of the plasmid and that fact that it exists in different forms (supercoiled, nicked and linear). As some strains lack a plasmid, the typeability is partial (Hartstein *et al.*, 1995; Weller, 2000).

1.8.2. Southern hybridisation

DNA is first digested with a restriction endonuclease and the restriction fragments are separated by gel electrophoresis and single-stranded DNA fragments are immobilised on to a nitrocellulose or nylon membrane. DNA molecules specific to certain sequences (probes) are hybridised to these fragments (Southern, 1992), thus reducing the number of fragments and aiding interpretation (Busch & Nitschko, 1999). Changing the specificity of the probes has led to the development of different typing techniques. Binary typing utilises up to 15 probes specific to MRSA. Based on hybridisation with each probe, each isolate is given a binary code. This technique is specific, easy to perform, reproducible and displays a high discriminatory power, but is time-consuming (van Leeuwen *et al.*, 1999; 2001). Ribotyping uses probes for the 16S rRNA gene which is present in multiple copies and positions on the

genome and is highly conserved within a species. It has been shown to be useful in outbreak situations and displays good reproducibility and discriminatory power, although it is less than that of other techniques such as PFGE (Bingen *et al.*, 1994; Prevost *et al.*, 1992; Tenover *et al.*, 1994; Yoshida *et al.*, 1997).

1.8.3. Pulsed-field gel electrophoresis (PFGE)

PFGE was developed by Schwartz and Cantor (1984) and is considered to be the present gold standard typing technique for S. *aureus* outbreaks (Bannerman *et al.*, 1995). It is able to rapidly detect accumulating variations and some movement of MGEs (Trindade *et al.*, 2003). The technique requires relatively unsheared DNA that is restricted with a 'rare-cutting' endonuclease. The endonuclease *Sma*I demonstrates the best performance for *S. aureus* (Carles-Nurit *et al.*, 1992; Ichiyama *et al.*, 1991; Prevost *et al.*, 1992). Thirty or less fragments ranging in size from 20 to 600 kb result in good discriminatory power (van Belkum *et al.*, 2007). Gel-based methods use a standard size ladder, thus limiting the accuracy of fragment sizing. Small differences in electrophoresis conditions can alter the banding pattern, thus making it difficult to standardise the technique between laboratories. This reduces the inter-laboratory reproducibility and makes data exchange and interpretation difficult; this has been overcome by the creation of international databases (van Belkum *et al.*, 2007). Data therefore needs careful analysis and quality control and standard interpretation of PFGE profiles has been proposed (Tenover *et al.*, 1995). Profiles differing by up to three bands are considered to be related whilst six or more band differences are thought to be exhibited between unrelated isolates. These criteria however only hold for isolates collected over a short period of time. The technique requires 2-4 days to obtain results and has high reagent and equipment costs. Despite this, PFGE has been useful in identifying and differentiating

MRSA (Carles-Nurit *et al.*, 1992). PFGE data in outbreak situations for example has helped identify the source and transmission events of MRSA (Ichiyama *et al.*, 1991).

1.8.4. Variable-number tandem-repeats (VNTRs)

Variable-number tandem-repeats (VNTRs, also known as minisatellites) are short nucleotide sequences that are repeated sequentially in the genome and are present in the same orientation. On average, the repeat unit lengths range from 9 to 65 bp and contain a short sequence known as a 'core' sequence (Jeffreys *et al.*, 1985; Wright, 1994). The distribution of these loci in the genome is not completely random and highly clustered in particular regions (Nakamura *et al.*, 1987). Polymorphisms are caused by the differing number of repeat units found at each locus. This variability is thought to result from unequal crossovers during chromosome replication, which occurs at a higher rate in minisatellites or slipped strand mispairing (van Belkum *et al.*, 1998). These are thought to be caused by problems with DNA repair during or after replication or with the enzyme DNA polymerase (van Belkum, 1999). VNTRs are thought to help the organism to adapt to environmental factors by accumulating factors providing a selective advantage (van Belkum *et al.*, 1998). Whether all VNTRs have a function within the genome is still debatable. VNTRs can be found within gene coding regions (Nakamura *et al.*, 1998) and variation probably can affect factors such as virulence in bacteria (van Belkum *et al.*, 1998).

VNTRs in *S. aureus* have been identified and are designated as staphylococcal interspersed repeat units (SIRUs). Seven SIRUs are common to all strains, although the number of repeats and their location vary between strains. These repeats contain similar flanking regions at the 3' and 5' end and lengths range from 48 to 159 bp. Three of the SIRUs are found multiple times in the genome. The sequences and numbers of repeat units differed

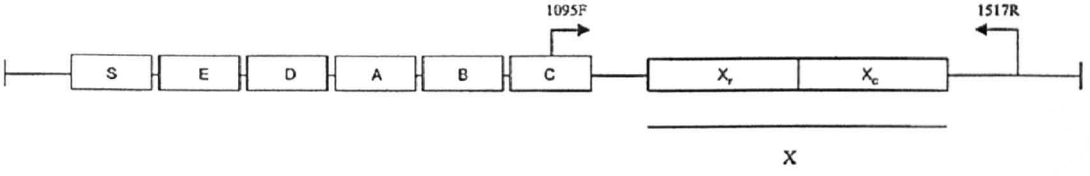
between each of the loci in regard to these SIRUs (Hardy *et al.*, 2004). SIRUs can discriminate both between EMRSA and endemic MRSA and results are concordant with MLST (Conceição *et al.*, 2009; Hardy *et al.*, 2004; 2006). PCR of multiple well characterised VNTRs can be utilised to differentiate between isolates. This method is technically simple and accessible to most healthcare facilities.. However, VNTR typing displays a discriminatory power lower than that of PFGE. In addition, the evolution of repetitive DNA is rapid, therefore care is needed to ensure related isolates in an outbreak are not discriminated. (Holmes *et al.*, 2010; van Belkum *et al.*, 2007).

1.8.5. Protein A (*spa*) typing

This technique is based on sequence determination of a VNTR unit found in the polymorphic X region at the 3' end of the *spa* locus (Figure 15). Protein A is found at the cell surface and is anchored into the cell wall. It is thought to play a role in the virulence of *S. aureus* (Shopsin *et al.*, 1999). This gene is under purifying selection (removal of deleterious variants therefore most mutations are synonymous) and therefore it is ideal for pandemic studies (Koreen *et al.*, 2004). The VNTR is composed of a 24 bp repeating unit (21 bp variant detected). Deletions, duplications and point mutations result in a PCR product ranging in size from 250 to 650 bp, which is detected by primers targeting the conserved flanking regions (Shopsin *et al.*, 1999). The data are largely concordant with MLST and although it has been shown to be useful in evolutionary and population genetic studies, this technique is limited in displaying chromosomal diversity (Koreen *et al.*, 2004). As *spa* typing is based on sequence data, it is unambiguous and data are stored on a *spa* database. More than 10, 000 *spa* types from over 90 countries have been deposited in the database (Ridom GmbH, 2012). Novel repeats can be submitted to the curated database for assignment to a *spa* type. The technique is applicable to all strains but is unable to detect chromosomal

changes outside this locus. However, *spa* typing shows good discriminatory power compared to bacteriophage typing (Shopsin *et al.*, 1999).

Figure 15. Structural composition of protein A gene



Boxes represent segments of the gene encoding for the signal sequence (S), the immunoglobulin G-binding regions (A-D), region homologous to A-D (E) and the carboxyl terminus (X). The terminus includes the short sequence repeat region (X_r) and the cell wall attachment sequence (X_c). The annealing sites for the forward (F) and reverse (R) primers are indicated by arrows and base pair position on the gene represented by numbers (obtained from Shopsin *et al.*, 1999).

1.8.6. Direct repeat units (*dru*) typing

This is a recent method based on the PCR and sequence determination of the *dru* locus found next to IS*431* in SCC*mec*. These sequences consist of 40 bp units and are found in the same location, regardless of the SCC*mec* type of the isolate. A study by Goering *et al.* (2008) utilised *dru* typing to track EMRSA-15 and EMRSA-16 isolates which present indistinguishable PFGE types. They showed the number of nucleotide changes in the units varied from one to seven; this variation in the sequence of the repeat units and the order of the units were designated a *dru* type. They distinguished 13 and 12 different *dru* types among the 47 EMRSA-15 and 57 EMRSA-16 isolates respectively. Other studies have shown that *dru* typing exhibit a Simpson index of diversity of 77.83 which was between that of *spa* typing

(66.9) and PFGE (81.34) (Shore *et al.*, 2010). Although these repeats are absent in a few MRSA strains, the sequences were found to be stable for use in strain determination.

1.8.7. Multilocus sequence typing (MLST)

The MLST scheme for *S. aureus* involves determining the sequence of approximately 450 bp of seven housekeeping genes encoding for proteins involved in central metabolism. These are carbamate kinase, shikimate dehydrogenase, glycerol kinase, guanylate kinase, phosphate acetyltransferase, triosephosphate isomerase and acetyl coenzyme A acetyltransferase (Enright *et al.*, 2000). Variations in these sequences are assigned as different alleles of the genes and are designated an allele number. The combination of the seven genes constitute the allelic profile of the strain, which constitutes an ST. These differences are not weighted and hence single or several sequence variations still result in a single change in allele and hence ST (Table 4). Up to 30 variants at each locus have been identified, which combines to a possible two million allelic profiles. This method is therefore highly discriminatory.

An online curated MLST database (http://www.mlst.net) has enabled the standardisation of MLST across the globe. The MLST website provides the eBURST software version 3 which displays the evolutionary relationship between isolates of a species. Closely related isolates are displayed as clusters known as CCs that aggregate around the 'founder' type which is the most closely related to all of the types within the cluster. Descendants of the founder are known as single-locus-variants (SLVs) if their allelic profile differs from the founding genotype at one allele and double-locus-variants (DLVs) if they differ at two. Each CC is thought to represent a different *S. aureus* lineage (Figure 16: Table 4) (Feil *et al.*, 2004). Ten major CCs of clinical relevance have been identified in *S. aureus* (Feil *et al.*, 2003). Those STs that are not assigned to a CC are known as singletons.

Table 4. Correlation between *S. aureus* characterisation methods and nomenclature

Characterisation method	Correlation
Lineage	Based on a single MLST* CC† which can be composed of multiple clones
CC	A lineage composed of closely related STs‡ based on MLST
ST	A subset of a CC, based on an allelic profile of seven housekeeping genes used for MLST
SCC*mec*§ types	Based on composition of genes present at a GI‖. Multiple SCC*mec* types are found in a lineage or CC
Clone	A subset of a lineage based on the combination of CC and SCC*mec* type
spa type	Based on protein A sequence, each *spa* type belongs to a subset of isolates in a single lineage

Displays nomenclature used in molecular techniques used to characterise isolates and correlations between the nomenclature terms.

***MLST, multilocus sequence type.**

†CC, clonal complex.

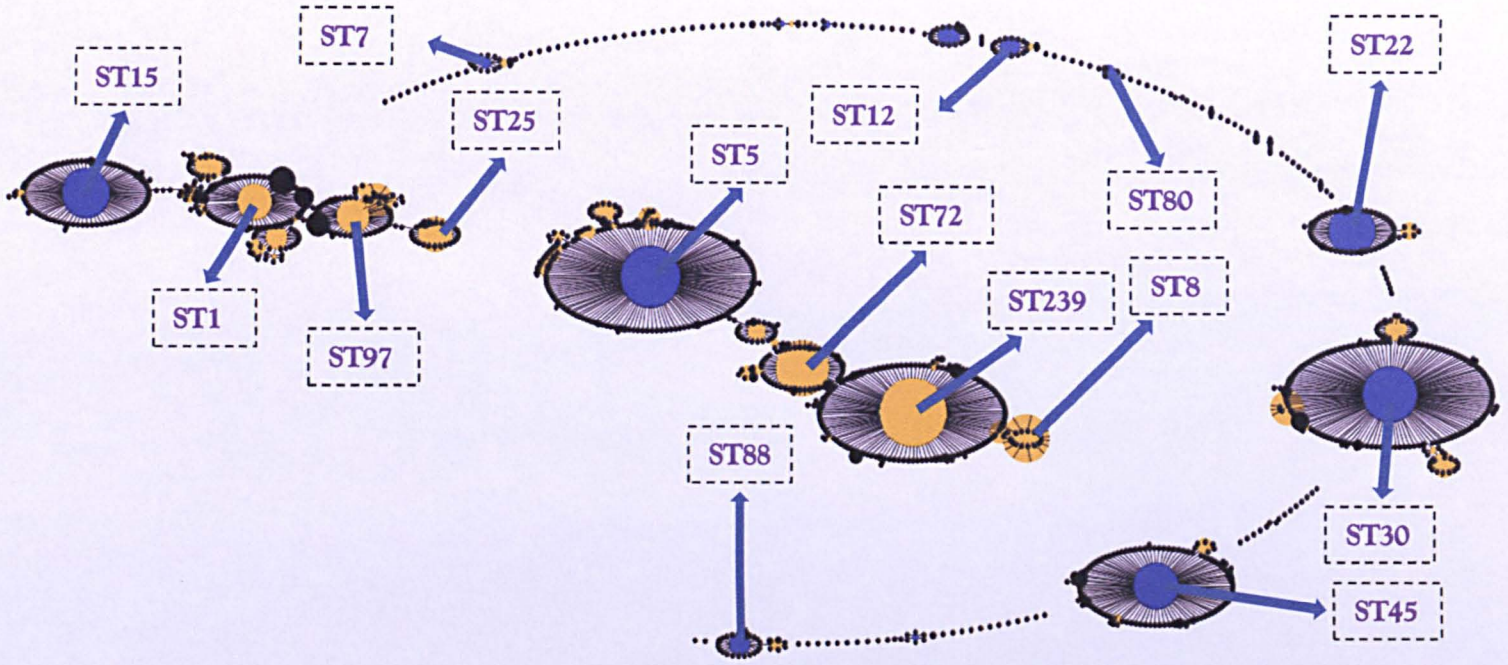
‡ST, sequence type.

§SCC*mec*, staphylococcus cassette chromosome *mec*.

‖GI, genomic island.

The variation in housekeeping genes is accumulated slowly. As isolates in an outbreak are likely to originate from a common ancestor, these isolates are unlikely to exhibit differences in these genes. Therefore MLST is not useful in epidemiology or outbreak studies. In addition, few lineages are found in one geographic area (Cockfield *et al.*, 2007). Despite its uses for studying the population structure, emergence and spread of clones, MLST remains expensive for routine hospital laboratories (van Belkum *et al.*, 2007). Nevertheless, MLST together with SCC*mec* typing has provided possible evolutionary models for MRSA.

Figure 16. eBURST diagram depicting population structure of the entire multilocus sequence typing (MLST) database for *S. aureus*



A 'population snapshot' of the entire MLST database (http://www.mlst.net) presented as an eBURST diagram (http://eburst.mlst.net) showing clusters of linked and unlinked STs (indicated by blue arrows and purple labels). The blue circles indicate the founder ST of a group and the yellow indicate subgroup founder STs. The size of the circle indicates the abundance of that ST in the database; larger circles indicate the presence of a greater number of isolates and vice versa.

1.8.8. SCC*mec* typing

Characterisation of MRSA clones by SCC*mec* typing provides insights into prevalent clones. Numerous assays for the detection of specific SCC*mec* types and subtypes have been described (Table 5). In addition to assigning isolates to MRSA clones, SCC*mec* typing helps our understanding of the virulence of MRSA. SCC*mec*-associated virulence factors for example are thought to play a role in the clinical outcome of bacteraemia (Ganga *et al.*, 2009). Other studies have shown HA-MRSA are significantly associated with SCC*mec* type II, *agr* group II and higher vancomycin MIC whilst CA-MRSA is associated with SCC*mec* type IV, *agr* group I and lower vancomycin MIC (Chua *et al.*, 2008; Moise *et al.*, 2009).

Table 5. Assays designed for SCC*mec* typing

SCC*mec* typing method	SCC*mec* target	Type of assay	Reference
1	Types I to V	Multiplex PCR	(Oliveira & de Lencastre, 2002)
2	Types I to V	Multiplex PCR	(Kondo *et al.*, 2007)
3	Types I to V	Multiplex PCR	(Zhang *et al.*, 2012)
4	Types I to V	Multiplex PCR	(Milheiriço *et al.*, 2007)
5	Subtypes IVa to IVh	Multiplex PCR	(Milheiriço *et al.*, 2007)
6	Types I to V	Multiplex PCR and DNA hybridisation	(Cai *et al.*, 2009)
7	Type VII	Multiplex PCR and real-time PCR	(Higuchi *et al.*, 2008)
8	Types I to VI and VIII	Multiplex real-time PCR	(Chen *et al.*, 2009)
9	Type VIII	Multiplex PCR	(McClure *et al.*, 2010)
10	Type XI	Multiplex PCR	(Stegger *et al.*, 2012a)

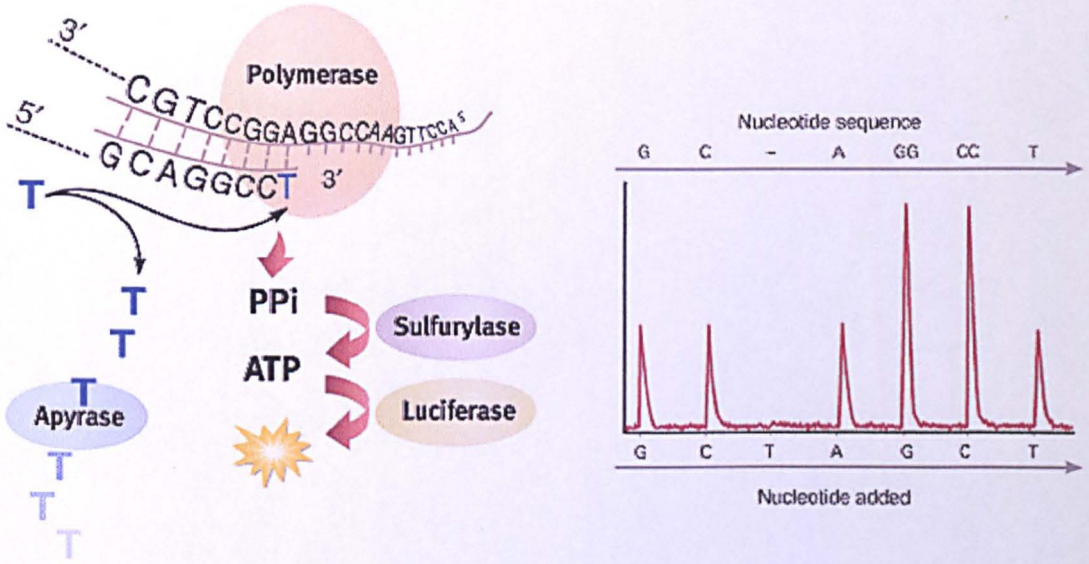
The table displays published studies that describe assays designed to detect specific SCC*mec* types and subtypes.

1.8.9. Pyrosequencing™

Pyrosequencing™ is an alternative sequencing-by-synthesis methodology to ddNTP termination technology. The method involves amplification of the template by PCR with a biotinylated primer. Biotinylated PCR products are captured on to streptavidin-coated magnetic beads to wash away primers, nucleotides and salts from the PCR reaction, which interfere with Pyrosequencing™ (Figure 17) (Gharizadeh *et al.*, 2002; Ronaghi *et al.*, 1998; Ronaghi, 2001). Pyrosequencing™ has a limited read length of around 100 bp and it is difficult to determine the number of nucleotides incorporated in regions composed of a series of identical nucleotides.

Pyrosequencing™ has been used in combination with Sanger sequence determination to obtain the whole-genome sequence (WGS) of a USA300 strain of *S. aureus* (Highlander *et al.*, 2007). It has been a useful tool in identifying *S. aureus* by targeting the bacterial 16S rRNA gene (Masoud *et al.*, 2012) and the *DEFB4* and *DEFB103* genes in humans, which have been investigated for associations with *S. aureus* bacteraemia (Fode *et al.*, 2012). It is also a useful tool for single-nucleotide-polymorphism (SNP) detection and analysis. Here, each sequence will produce a distinct Pyrogram™ enabling the easy assignment of alleles (Ahmadian *et al.*, 2000; Fakhrai-Rad *et al.*, 2002). This methodology is accurate, rapid and results can be obtained in real-time.

Figure 17. Principle of Pyrosequencing™



The figure depicts the principle of Pyrosequencing™ and an example of the output Pyrogram™. During Pyrosequencing™, each nucleotide is added into the reaction in succession. Incorporation of a nucleotide into the nucleic acid chain is catalysed by the polymerase. This step releases pyrophosphate (PPi) which is converted to ATP by sulfurylase present in the reaction. The ATP is used by luciferase to produce light which is detected by a camera. The intensity of the light is proportional to the number of nucleotides added, enabling sequence determination. Apyrase degrades unincorporated nucleotides. Incorporation of a single nucleotide is indicated by the lower height peaks whilst incorporation of two nucleotides is indicated by a double peak height (obtained from England & Pettersson, 2005).

1.8.10. Microarrays

Microarrays for *S. aureus* genes have been developed from up to seven WGSs (Witney *et al.*, 2005). DNA samples hybridise to complementary labelled sequences on the array, identifying the presence of this sequence in the sample. This enables comparative genomic

studies as the presence and absence of genes can be identified in isolates. In addition, as messenger RNA can be reverse transcribed into complementary DNA, transcription studies can also be performed (Holden *et al.*, 2008).

Comparative studies of *S. aureus* using microarrays have revealed genetic variation between isolates causing similar clinical symptoms. A proportion of this variation was due to large-scale deletions, some of which were found at loci with MGEs (Fitzgerald *et al.*, 2001). Other studies have also identified variations in MGEs; however this variation was not correlated with isolates causing invasive disease. This study therefore concluded colonising isolates (presence of bacteria on a body surface without displaying clinical symptoms of an infection) and those causing disease can be genetically similar (Lindsay *et al.*, 2006). Microarray studies, along with MLST, have also revealed the population structure of human *S. aureus* which belong to ten dominant lineages and several minor lineages. Excluding core genes present in every lineage, a combination of genes amongst the variable region of the genome are unique to each lineage. These genes encode a variety of toxins, proteases and superantigens, with a large proportion encoding for surface proteins. These unique combinations of proteins in each lineage may increase our understanding of the predominance of one or two lineages in a geographical location (Cockfield *et al.*, 2007; Feil *et al.*, 2003; Lindsay *et al.*, 2006). Microarrays have also revealed some MGEs are lineage-specific whilst others are variable within a lineage or are found between multiple lineages.

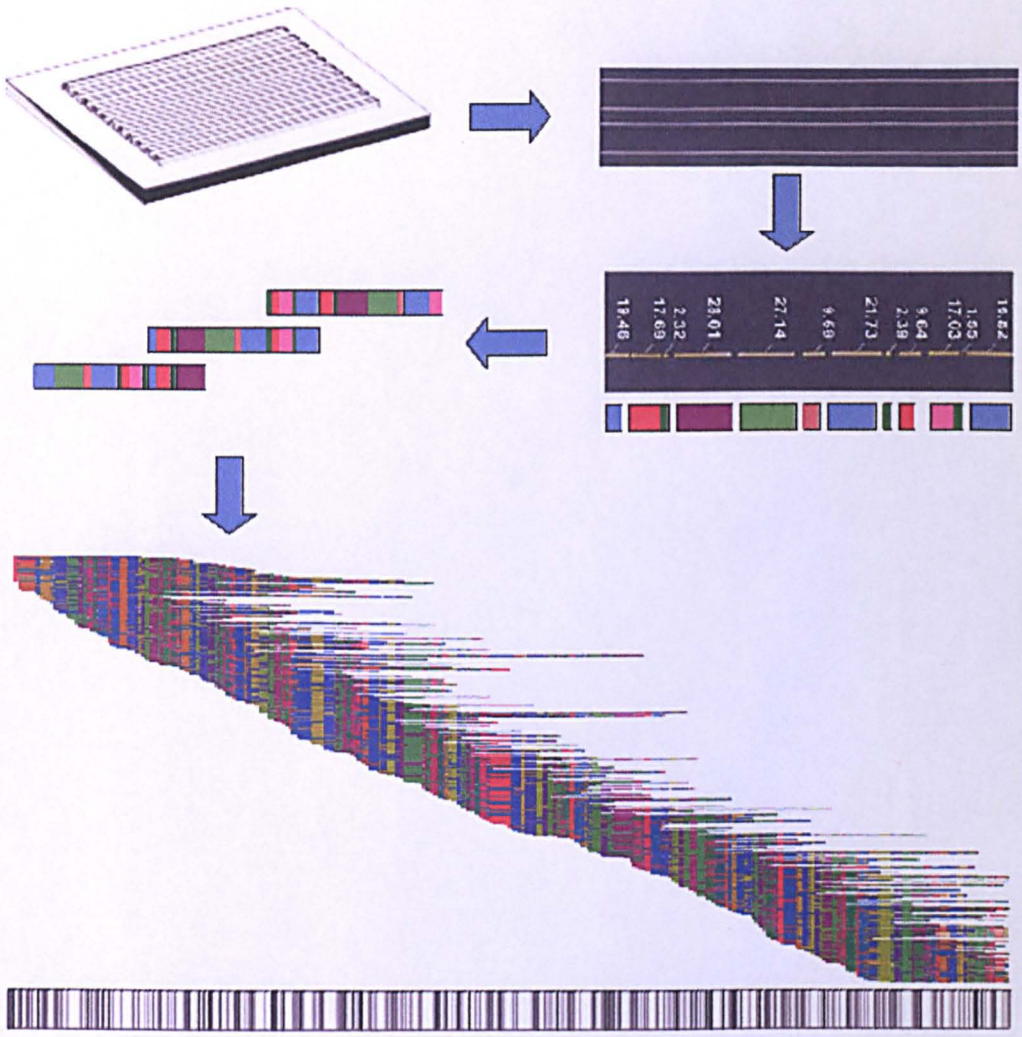
1.8.11. Whole-genome mapping (WGM)

Whole-genome mapping (WGM, formerly optical mapping) is a technique that utilises a restriction endonuclease to produce an ordered whole-genome restriction map (optical map) with high-resolution. The technique involves gentle lysis of cells to isolate high molecular

weight DNA. The fragmented DNA (approximately 200 kb) is restricted with a rare-cutting endonuclease and optical maps are generated (Figure 18). These can be aligned in WGM software to perform genome comparisons. In addition, *in silico* maps can be generated from National Center for Biotechnology Information (NCBI) database sequences to enable comparative genomics and strain typing.

Comparative genomics of strain maps enables the detection of insertions, deletions and large-scale rearrangements. It has been successfully utilised to identify a large inversion in an *S. aureus* strain (Shukla *et al.*, 2009). The alignment of maps can also be used for epidemiological studies and outbreaks and can differentiate between strains. In *S. aureus*, optical maps enable the differentiation between strains harbouring different SCC*mec* cassettes. Similarity clustering data of *S. aureus* strains was concordant with those obtained by PFGE (Shukla *et al.*, 2012). In addition, MRSA-specific motifs associated with virulence were identified. To date, optical maps have helped close gaps in multiple *S. aureus* genomes (Chua *et al.*, 2010; Huang *et al.*, 2012; Stegger *et al.*, 2012b). At present, the cost and rapidity of optical maps do not make it feasible for routine diagnostics.

Figure 18. WGM protocol



For whole-genome mapping (WGM) DNA is captured along parallel arrays on a microfluidics device by capillary action and held electrostatically to the surface. The DNA molecules are restricted with a rare-cutting endonuclease and fragments are fluorescently labelled. Software scans the restricted molecules in each lane and assigns the fragments a size and order; these generate a digital single molecule restriction map. The single molecule restriction maps are aligned to produce a map covering the entire genome. This alignment provides deep coverage which improves accuracy and allows for incomplete digestion or random breaks in the DNA. A consensus map is generated from this alignment (obtained from Fitzpatrick & Burggrave, 2009).

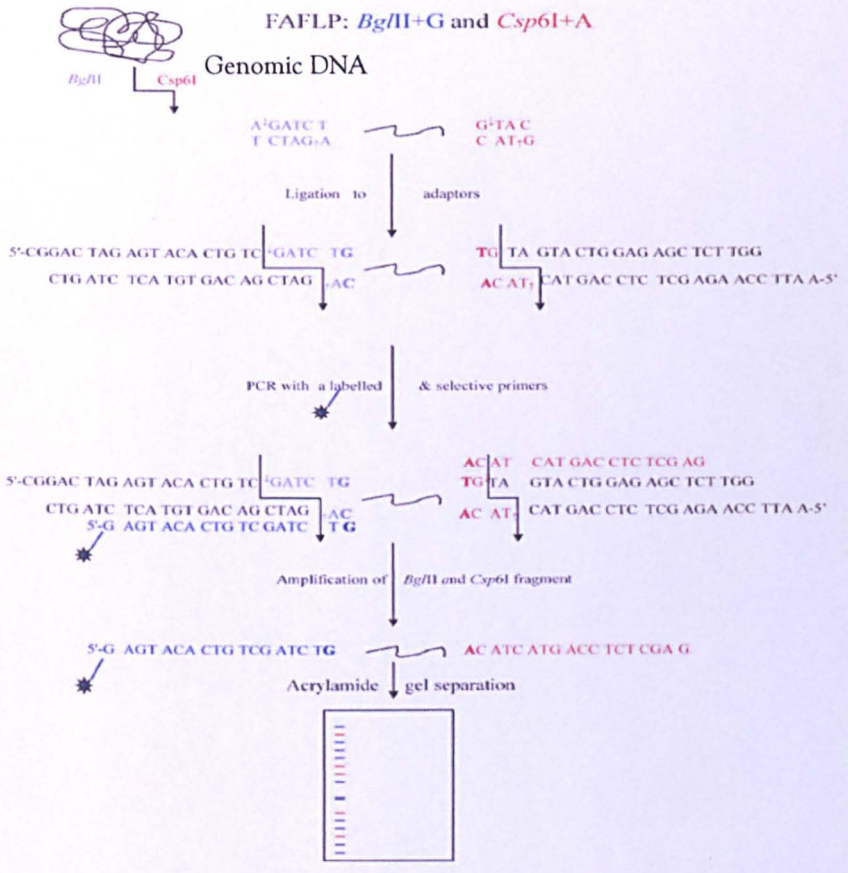
1.8.12. Next-generation sequencing (NGS)

WGSs have been utilised for the development of microarrays which have been useful in comparative genomics. Since the advent of NGS technologies, the cost and time to sequence a genome have fallen dramatically. These improvements could lead to NGS being utilised in routine laboratories in the future. In a clinical setting, NGS can identify strains associated with an outbreak and distinguish them from background strains. Furthermore, NGS can identify transmission events associated with the outbreak, including those within the community setting. Such timely insights can identify the source of an outbreak and enable the implementation of affective infection control measures. However, use of NGS as a routine diagnostic tool still requires the development of WGS databases for comparisons and an automated analysis method (Harris *et al.*, 2013; Köser *et al.*, 2012).

1.8.13. Fluorescent amplified fragment length polymorphism (FAFLP)

Fluorescent amplified fragment length polymorphism (FAFLP) is a PCR-based DNA fingerprinting technique which detects polymorphisms in primer annealing sites and restriction endonuclease recognition sequences. In addition, it also detects insertions and deletions within core conserved regions and variable genomic regions (Vos *et al.*, 1995). FAFLP utilises two restriction endonucleases and ligation of restricted fragments to double-stranded adaptors (Figure 19). PCR using selective or non-selective primers, one of which is labelled with a fluorescent dye, results in amplification of a subset of fragments. Selective primer sequences comprise additional nucleotides (up to four) on the 3' end of the sequence; each additional nucleotide (+X) reduces the number of fragments amplified by four-fold.

Figure 19. Diagram of fluorescent amplified fragment length polymorphism (FAFLP)



Genomic DNA is restricted using two endonucleases, a 'rare-cutter' (blue: *Bgl*II) which recognises a 6 bp sequence and a 'frequent-cutter' (red: *Csp*6I) which recognises a 4 bp sequence. The majority of resultant fragments are restricted at both ends with the frequent-cutter, a minority are restricted with the rare-cutter and a moderate number are restricted with both endonucleases. The latter are the fragments of interest. Double-stranded adaptor sequences composed of a core sequence and a sequence complementary to the restriction endonuclease recognition sequence are ligated to the ends of the restricted fragments. PCR is performed using primers with sequences complementary to the adaptor sequences. One of the primers is fluorescently labelled at the 5' end (blue sequence). The amplified fragments are run on an automated capillary sequencer along with a size standard in each lane (obtained from Desai, 1999).

Fragments are run on an automated capillary sequencer along with an internal size standard labelled with a different fluorescent dye. This enables accurate sizing of the fragments to within ±0.5 bp, thus improving the reproducibility between laboratories (Arnold *et al.*, 1999; Desai *et al.*, 1998). FAFLP analysis can be performed *in silico* on WGSs using software and online tools to determine the sequences of fragments of interest (Bikandi *et al.*, 2004; Underwood, 2008).

Utilising two endonucleases ensures small fragments are generated which can be resolved well by capillary gel electrophoresis. Simultaneously, use of a rare-cutting endonuclease limits the number of fragments generated. Utilising two endonucleases also enables one strand of the fragment to be fluorescently labelled. As only one strand is detected, the analysis is not compromised by unequal mobility of the two strands. The use of two endonucleases provides greater flexibility in the number of fragments and profiles generated (Vos *et al.*, 1995).

FAFLP is a robust and high-resolution methodology used for the analysis of genetic diversity within bacterial genomes, without prior knowledge of the genome sequence. This technique covers a large proportion of the genome and exhibits a high discriminatory power. Altering the primers and restriction endonucleases can increase or decrease the discriminatory power, as different regions of the genome are sampled. As the amplified fragments (AFs) are selected randomly, the technique is representative of the whole genome. It has drawbacks in that variations outside of the restriction enzyme sites are not detected, although it can, to an extent, detect transposition events (Vela *et al.*, 2011). The results are reproducible within the same laboratory if the same capillary sequencer is used. FAFLP can distinguish between isolates that are indistinguishable from one another by other methods such as restriction

fragment length polymorphism, PFGE and serotyping (Desai *et al.*, 1999; Hookey *et al.*, 1999). It displays greater reproducibility than PFGE but less so than MLST for *S. aureus* (Melles *et al.*, 2007). Isolate clusters based on VNTRs however, are similar to those obtained by AFLP (Melles *et al.*, 2009).

The application of FAFLP has enabled the discrimination of genetically related bacterial strains and has helped in outbreak analysis and epidemiological studies for varied bacterial genera (Desai *et al.*, 1998; Grady *et al.*, 1999; Janssen *et al.*, 1996; Kober *et al.*, 2011). In addition, FAFLP has been applied to population structure studies and bacterial taxonomy (Janssen *et al.*, 1996; Melles *et al.*, 2008) and can detect transposition (Vela *et al.*, 2011). FAFLP has been adapted to target IS*6110* in *Mycobacterium tuberculosis* to provide insights into its evolution. This involves the identification and amplification of the IS present in variable copy numbers and at variable positions in the genome between different strains (Thorne *et al.*, 2007). Epidemiologically unrelated but genetically related isolates can be distinguished using this method (Borrell *et al.*, 2009). In addition, FAFLP has assisted the identification of hypervirulent lineages (Goulding *et al.*, 2000). Isolates in the same FAFLP cluster of organisms share a number of common AFs but also display additional differential fragments highlighting the discriminatory power of FAFLP (Desai *et al.*, 2001).

FAFLP studies on *S. aureus* have enabled differentiation between various bacteriophage types and has revealed the genetic heterogeneity within EMRSA-15 (Grady *et al.*, 1999; 2000). FAFLP cluster analysis does not separate MRSA from MSSA isolates but AFs exclusive to MRSA and MSSA have been identified. Comparative studies of invasive and colonising clinical isolates using amplified fragment length polymorphism (AFLP) have identified that both types of isolates belong to similar genetic backgrounds as multiresistant isolates.

However, invasive isolates are overrepresented in some genetic clusters (Melles *et al.*, 2004). This correlates with the later findings of Lindsay *et al.* (2006) who found no association of lineages with the severity of infection or with any single virulence or pathogenicity factor. AFLP can highlight the genetic heterogeneity within major lineages of *S. aureus*. AFLP genetic clusters and sub-clusters correlate with MLST CCs. Previous studies have shown that whilst isolates of the same lineage shared a greater number of fragments compared to isolates from different lineages, unique AFLP markers could be identified amongst isolates of a specific cluster. AFLP has been applied to short-term epidemiological studies and population structure studies of *S. aureus* (Melles *et al.*, 2004; 2007; 2009).

S. aureus typing techniques that target single or multiple loci detect genetic variation at these loci. Inferences based on this variation can be made about the genetic diversity amongst this species. Molecular tools may target core regions of the genome (MLST), variable regions (VNTR and SCC*mec* typing) or may target regions throughout the genome (microarrays, WGM and FAFLP). Each molecular method displays a number of advantages and disadvantages and the choice of molecular method will depend on the focus of the study, whether that be in the investigation of population structure, an outbreak or genetic diversity. Studies investigating the genetic diversity amongst a species may utilise molecular methods which target a large number of loci, core and variable regions of the genome and regions throughout the genome. This would present an overview of the genetic variation present across the genome. Thus molecular tools such as FAFLP and WGM present an advantage as they require no prior sequence knowledge.

1.9. Aims and objectives of this study

The aim of this study was to investigate the genetic diversity between and within predominant hospital-associated MRSA lineages in the United Kingdom. This study hypothesised that distinct genetic differences are present between successful and unsuccessful lineages of MRSA. Regions of genetic variation were identified between and within lineages. Sequence data of regions of genetic heterogeneity between major MRSA lineages was mapped to the genome and linked to their cell functions. This study explored the role of these cell functions in the emergence and prevalence (or spread) of dominant genetic lineages. These mechanisms may confer a fitness advantage by altering the virulence or pathogenicity between lineages.

CHAPTER 2

2. Materials and Methods

2.1. Bacterial culture collection

A total of 302 S. *aureus* isolates were included in this study; arbitrary numbers, DBHT 1 to DBHT 302, were given to the isolates. The isolates were obtained from five collections. The five collections included 50 isolates from SRL (henceforth referred as SRL collection), six strains obtained from the National Collection of Type Cultures were used as reference strains (henceforth referred as Reference collection), 34 isolates obtained from Glasgow, Public Health Scotland (henceforth referred as Glasgow collection), 202 isolates from the 2009 British Society for Antimicrobial Chemotherapy bacteraemia resistance surveillance programme (henceforth referred as BSAC collection) (British Society for Antimicrobial Chemotherapy, 2010) and 10 isolates obtained from the Clinical Microbiology and Public Health Laboratory, PHE, Cambridge (henceforth referred as CMPHL collection).

The first collection of 50 isolates from SRL was sent to the laboratory for identification and characterisation during 2007 to 2009 from geographically diverse centres throughout England and Wales. Bacteriophage typing and PFGE were performed by SRL. Bacteriophage typing identified four isolates as EMRSA-15, five belonged to bacteriophage type A, one was a wide-spread sporadic strain, 11 displayed previously unrecognised distinct bacteriophage patterns, nine were non-typeable and no data were available for the remaining 20 isolates. PFGE identified 15 isolates as EMRSA-15, four displayed pattern A, pattern B and C were displayed by one isolate each, five displayed distinct PFGE patterns and no data were available for the remaining 24 isolates. Of the 50 isolates, 25 were obtained from eight hospitals and no data on the origin of the remaining 25 isolates was available. Eight isolates from a single hospital were obtained from a single patient. Bacteriophage typing results were

available for seven of the eight isolates; two were identified as EMRSA-15 whilst five were non-typeable. PFGE results were available for seven of the eight isolates; six were identified as EMRSA-15 whilst one exhibited pattern A. Each of the four isolates from a single ward displayed a distinct bacteriophage pattern.

WGSs for all of the six Reference collection strains were available on the NCBI database (http://www.ncbi.nlm.nih.gov) (Table 6). STs for these strains have been previously described, based on their WGS (Saunders *et al.*, 2004). Of the six strains, DBHT 54 and DBHT 59 belonged to ST5 whilst DBHT 79 and DBHT 84 were assigned to ST1. The two remaining strains, DBHT 51 and DBHT 75, belonged to ST36 and ST250 respectively.

Table 6. Characteristics of six strains from the Reference collection

Isolate number	Isolation year	Antimicrobial resistance	Strain	Comments	Accession numbers*
DBHT 51	ND†	MRSA	MRSA252	HA‡	NC_002952
DBHT 54	1982	MRSA	N315	HA	NC_002745
DBHT 59	1996	VISA	Mu50	HA	NC_002758
DBHT 75	ND	MRSA	COL	ND	NC_002951
DBHT 79	1998	MRSA	MW2	High virulence CA§	NC_003923
DBHT 84	ND	MSSA	MSSA476	High virulence CA	NC_002953

Displays year of isolation, meticillin or vancomycin resistance, and origin of six reference strains with whole-genome sequence (WGS) data.

***Accession numbers obtained from the NCBI database for WGS data for each strain.**

†ND, no data.

‡HA, hospital-associated.

§CA, community-associated.

The Glasgow collection of 34 isolates was previously characterised using bacteriophage typing, PFGE, oxacillin MICs, and *mecA* and *mupA* PCRs at SRL, PHE (Appendix I). Bacteriophage typing was performed on 27 isolates; five were non-typeable, two belonged to the same distinct type and the remaining 20 isolates displayed a unique bacteriophage pattern each. Based on PFGE results, three and four isolates were assigned to EMRSA-15 and -16 respectively. Of the remaining 27 isolates, 26 isolates were assigned to 26 unique PFGE profiles each and no datum was available for one isolate. Oxacillin MIC was performed on 32 isolates. The majority of isolates were resistant (n=23) or displayed borderline resistance (n=7); the remaining two isolates were susceptible to oxacillin. Twenty-three of the 27 isolates tested positive for the *mecA* PCR and the remaining four isolates were negative. Of the 18 isolates analysed for the *mupA* gene, the majority (n=15) were negative.

The BSAC collection consisted of 99 MRSA and 103 MSSA isolates. The MRSA and MSSA isolates were contributed from a total of 25 and 24 centres respectively throughout England and Ireland (Appendix II) (Reynolds *et al.*, 2008). Each centre was able to contribute up to 20 consecutive isolates. The MICs of six antimicrobials, including oxacillin and vancomycin, were determined by the BSAC agar dilution method (Appendix II). PCRs for the *mecA* and *mupA* genes were performed at the Antimicrobial Resistance and Healthcare Associated Infections Reference Unit (PHE, London; Appendix II) (Andrews, 2001).

Of the 202 BSAC isolates, MLST data (see materials and methods: section 2.4.1) from a subset of 71 isolates was analysed further with antimicrobial MICs (Appendix II) and SCC*mec* data (see materials and methods: section 2.4.2). These 71 isolates included one representative of a specific lineage selected randomly from each centre to ensure each isolate

was independent. The percentage resistance amongst isolates of each lineage was calculated using Stata software version 12 (StataCorp LP, USA) for each of the six antimicrobials along with the 95 % CI. The percentage of isolates of a specific lineage belonging to each of the SCC*mec* types amongst isolates of each lineage was also calculated along with the 95 % CI.

The CMPHL collection of 10 isolates consisted of six isolates from an outbreak in the intensive care unit (ICU) and were collected in December 2009. Of these, five isolates were EMRSA-15 by bacteriophage typing and/or PFGE (Table 7). Antimicrobial susceptibility testing was performed on the six isolates from ICU patients for eight antimicrobials (Appendix III). The remaining four isolates were collected over June to December 2009.

Table 7. Characteristics of the 10 *S. aureus* isolates obtained from the Clinical Microbiology and Public Health Laboratory

Isolate number	Isolation year	Bacteriophage typing result	PFGE result	Comments
DBHT 293	2009	EMRSA-15*	EMRSA-15	ICU† outbreak
DBHT 294	2009	EMRSA-15	EMRSA-15	ICU outbreak
DBHT 295	2009	ND‡	ND	
DBHT 296	2009	ND	ND	ICU outbreak
DBHT 297	2009	ND	ND	
DBHT 298	2009	ND	ND	
DBHT 299	2009	ND	ND	
DBHT 300	2009	ND	EMRSA-15	ICU outbreak
DBHT 301	2009	EMRSA-15	EMRSA-15	ICU outbreak
DBHT 302	2009	EMRSA-15	EMRSA-15	ICU outbreak

The year and ward of isolation, bacteriophage and PFGE type of each isolate is displayed.

***EMRSA, Epidemic MRSA.**

†ICU, intensive care unit.

‡ND, no data.

The SRL collection provided multiple isolates from the same centre including those obtained from a single patient. The SRL collection was utilised as a test panel of isolates to develop and optimise a novel FAFLP method. This collection was analysed using FAFLP to identify whether FAFLP could identify genetic heterogeneity within isolates from a single patient. The Reference collection facilitated *in silico* analysis on strains and a direct comparison with experimental data for the same strains. The Glasgow collection provided a well-characterised isolate collection from Scotland to ensure the study represented isolates from different regions of the UK. The BSAC collection provided a well-characterised set of isolates to assess the genetic diversity between lineages using FAFLP. The CMPHL collection provided a collection of invasive and carriage isolates associated with an outbreak. FAFLP was used to assess whether FAFLP could identify genetic heterogeneity within these isolates.

2.2. Bacterial cultures and DNA extraction

Isolates were inoculated on Columbia agar plates supplemented with horse blood (CBA) (Oxoid, Thermo Scientific, UK), incubated overnight at 37°C in air and were stored on Microbank™ preservation beads (Pro-lab diagnostics, UK) at -80°C. Two MagNA Pure systems (Roche, UK), a compact and a high-throughput system, were utilised for DNA extraction.

2.2.1. MagNA Pure compact

Genomic DNA from a subset of isolates was extracted using the MagNA Pure compact system (Roche, UK). Cell lysis was performed by selecting up to five single colonies from 24 h cultures with a 1 µl loop and re-suspending in 180 µl of lysis buffer. Lysis buffer was composed of 2x Tris-EDTA buffer, 1.2 % Triton™ X-100, 0.3 mg/ml lysozyme, 0.03 mg/ml

lysostaphin (Sigma-Aldrich, UK) and nuclease free water (hereafter referred to only as water) (Promega, UK). The tubes were incubated for 1-2 h at 37°C and the final volume was made up to 400 µl with water. DNA was extracted using the MagNA Pure compact system and the MagNA Pure compact nucleic acid isolation kit I (Roche, UK) according to the manufacturer's recommendations.

2.2.2. MagNA Pure LC

Genomic DNA for the remaining isolates was extracted using the MagNA Pure LC (Roche, UK). Isolates from CBA plates were re-suspended in lysis buffer, as instructed in the MagNA Pure LC DNA isolation kit III (Bacteria, Fungi) (Roche, UK). The inoculated lysis buffer was incubated at 65°C for 4 h. DNA was extracted using the MagNA Pure LC system and MagNA Pure LC DNA isolation kit III according to the manufacturer's recommendations. With both extractions, DNA was eluted into 100 µl of elution buffer. Stock DNA solution was quantified using the NanoDrop® 8-sample spectrophotometer ND-8000 (Labtech international, UK). A 1:10 diluted stock solution was used for further experiments, unless otherwise stated.

2.3. *S. aureus* identification

2.3.1. 16S ribosomal RNA (rRNA) gene amplification and sequence determination

A subset of isolates was identified to species level by 16S rRNA gene PCR and sequence determination. Confirmatory identification was required for those isolates from a single patient which were not identical based on MLST and SCC*mec* typing. PCR was performed under standard conditions using an annealing temperature of 56°C (described below) using

the universal 16S rRNA gene primers ANTIF and 1392R (MWG Eurofins, UK) (Appendix IV). Sequences were determined using the standard sequence determination protocol (described below), and using 16S rRNA gene sequencing primer 357F or 3R (MWG Eurofins) (Appendix IV). The resultant chromatograms of the 16S rRNA gene sequences were assembled into a contig using BioEdit software version 7.0.9.0. The consensus sequence was searched against the NCBI database using Basic Local Alignment Search Tool (BLAST). Isolates were assigned to a genus and species based on the consensus of the top search results and percentage homology.

2.3.1.1. Standard PCR conditions

PCR was performed in a 50 µl PCR reaction containing 1x PCR reaction buffer, 2.5 mM $MgCl_2$, 0.2 mM each deoxynucleotide triphosphate (dNTP), 1.25 U *Taq* DNA polymerase recombinant (Life Technologies, UK), 0.2 µM of each primer and 50 ng of DNA. Thermocycling was performed on GeneAmp® PCR system 9700 (Life Technologies) and cycling conditions used were as follows: 95°C for 2 min; followed by 35 cycles of 95°C for 45 s, X°C for 45 s (where X is an annealing temperature specific to each primer pair) and 72°C for 1 min and a final extension at 72°C for 5 min with a hold cycle at 4°C. PCR products (5 µl) were confirmed on a 2 % agarose gel (Bioline, UK) pre-stained with 1x GelRed™ nucleic acid stain (Biotium, UK) and separated by electrophoresis at 80 V for 2 h. Products were run with 1 µl of 1x BlueJuice™ gel loading buffer and sized using 1 µl of 100 bp ladder (Life Technologies). DNA within gels was visualised under ultraviolet light using the Gel Doc 2000 (Bio-Rad, UK).

2.3.1.2. Standard sequence determination protocol

PCR reaction mixtures were 'cleaned up' to remove unincorporated dNTPs, primers and other salts or contaminants using the Agencourt® AMPure® reagent and the Biomek NX^P robot (Beckman Coulter, UK) following the manufacturer's protocol. Bidirectional sequencing was performed using a forward and reverse primer. Sequencing PCR reactions were performed according to the ddNTP termination method by Sanger (Sanger *et al.*, 1977) using the BigDye® Terminator version 1.1 cycle sequencing chemistry (Life Technologies). Each 10 µl sequencing reaction contained 1 µM of sequencing primer, 1.5 µl of 5x sequencing buffer, 1.0 µl ready reaction dye-mix, 5.5 µl of water and 1 µl of PCR template. The thermocycling conditions used were as follows: 96°C for 1 min, followed by 25 cycles of 96°C for 10 s, 50°C for 5 s, 60°C for 1.15 min and a hold of 4°C. Sequencing reaction mixtures were cleaned up using the Agencourt® cleanSEQ® reagent (Beckman Coulter, UK) and the Biomek NX^P robot following the manufacturer's protocol. This process eliminates dye-terminators and other sequencing reaction components such as primers, salts and contaminants. Sequences were then analysed using the ABI 3730 Genetic Analyzer (Life Technologies). Performance Optimized Polymer (POP-7™) (Life Technologies) was used for capillary electrophoresis at 13.4 kV with a 3 s injection time and 1.5 kV injection voltage. A 48-capillary array with a length of 50 cm was utilised. Negative sequencing controls with no PCR template were included in each run. Positive sequencing control included an *E. coli* plasmid vector, plasmid DNA pGEM® 3Zf (+), which was amplified using the M13 (-21) sequencing primer (Life Technologies). Samples with ambiguous bases in the forward and reverse chromatograms were repeated.

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

Thirty-two isolates chosen at random from amongst the SRL, Reference, Glasgow and BSAC collections were identified to the species level by MALDI and spectra differences were analysed, as previously described (Wolters *et al.*, 2011). Protein extraction was performed using an ethanol and formic acid protocol developed by Bruker Daltonics (Coventry, UK). One bacterial colony was resuspended in 300 µl of water, mixed well and 900 µl of absolute ethanol was added. The solution was centrifuged at 12 000 x g for 2 min and the supernatant was decanted; this process was repeated to form a pellet. Fifty µl of 70 % formic acid was added to the pellet and mixed thoroughly. To this solution, 50 µl of acetonitrile was added and centrifuged for 2 min at 12 000 x g. One µl of the supernatant was spotted three times on an MSP 96 polished steel target plate. Samples were air dried and 1 µl of saturated α-cyano-4-hydroxycinnamic acid (HCCA) matrix solution (matrix crystals were dissolved in water containing 50 % acetonitrile and 2.5 % trifluoroacetic acid, TFA) was spotted onto each sample and left to dry in air. A bacterial test standard (BTS) (Bruker Daltonics) composed of *E. coli K12* protein extract and two additional higher mass range proteins (RNAse A and myoglobin) were used for calibration of the mass spectrometer. The BTS covered a mass range from 4 to 17 kDa and was used as a positive control; HCCA matrix was used as a negative control. The controls and test samples were spotted in duplicate on to the target plate. All reagents were obtained from Sigma Aldrich unless otherwise stated.

The target plate was cleaned using a protocol developed by Bruker Daltonics. In brief, the target plate was overlaid with 70 % ethanol and incubated for 5 min. After rinsing the target plate with water, it was wiped with 70 % ethanol and a Kimwipe®. The target plate was

rinsed with water and wiped dry with another Kimwipe®. After overlaying the target plate with 80 % TFA in a fume hood, it was wiped vigorously with another Kimwipe®. The target plate was rinsed with deionised water, wiped, left to dry for 15 min and stored in a box.

Samples were analysed using the Bruker microflex™ mass spectrometer software. The samples were subjected to a laser frequency of 60 Hz; a mean spectrum was generated from 240 spectra for each spot. The spectra detected peaks between 1960 to 20137 m/z; proteins within this range are mainly ribosomal proteins. The spectra were compared using Biotyper software version 3.0 (Bruker Daltonics) against a database curated by Bruker Daltonics, which contains quality controlled spectra for >2000 microbial species. Peaks that match against the spectra in the database are given a positive value whilst those that are different are assigned negative values. A score is given to the top matches against the database and depending on the score, a species and/or genus identification is obtained. The isolates were analysed by MALDI in replicate at both CMPHL and at Colindale, PHE, London.

The mean spectrum for each isolate was analysed manually using FlexAnalysis software version 3.0 (Bruker Daltonics, UK). Each spectrum was analysed for the presence or absence of 13 m/z peaks; the combination of the presence and absence of these 13 peaks constituted the MALDI type. Based on this analysis, the isolates were assigned to one of 15 MALDI-types. These corresponded to one of the following CCs: CC5, 8, 22, 30 or 45 (Wolters *et al.*, 2011). Isolates that did not belong to any of the 15 types were presumed to belong to a CC not identifiable by the presence or absence of these 13 peaks.

2.4. Molecular tools for characterising strains

2.4.1. Multilocus sequence typing (MLST)

MLST of *S. aureus* was performed on all 302 isolates. PCR was performed under standard conditions (see materials and methods: section 2.3.1) using an annealing temperature of 55°C (Enright *et al.*, 2000). Sequences from the seven housekeeping genes *arc*C, *aro*E, *glp*F, *gmk*, *pta*, *tpi* and *yqiL* were amplified using seven primer pairs (Appendix IV). Sequencing reactions using the PCR primers were performed using the standard sequence determination protocol described (see materials and methods: section 2.3.1). Chromatograms were assembled into contigs using BioNumerics software version 6.1 (Applied Maths, Belgium). The sequence of each amplicon was assigned to an allele by comparison with the MLST database using BioNumerics (Platt *et al.*, 2006). Isolates were assigned to a CC based on an eBURST diagram produced from the entire MLST database (Feil *et al.*, 2004). The eBURST diagram was generated using the definition of a group as those STs that shared five or more alleles with one another.

Novel sequences that did not correspond to a known allele were submitted to the MLST database for confirmation. Sequences submitted to the database require good quality forward and reverse chromatograms which span the entire target region. Chromatograms for those isolates that displayed a novel sequence at one of the six loci (*arc*C, *aro*E, *glp*F, *pta*, *tpi* or *yqiL*) were submitted to the *S. aureus* MLST database curator for allele assignment. Redesigning the existing *gmk* reverse primer was required to enable submission of chromatograms for those isolates displaying a novel allele at the *gmk* locus. Sequencing reactions start downstream of the primer sequence. However, as the *gmk* reverse primer is

complementary to a sequence within the target region, chromatograms spanning the entire *gmk* region could not be obtained using the original primer. Fifteen *gmk* gene sequences were downloaded from *S. aureus* WGSs from the NCBI database (Appendix V: Group A) into BioEdit software. A consensus sequence was generated with gaps at polymorphic sites. Primers were designed utilising Primer3 software (Rozen & Skaletsky, 1999) and the consensus sequence. The *gmk* primers thus designed and used for amplification were: 5' GGATAATGAAAAAGGATTGTTAATCG 3' (*gmk*F_F2) and 5' CGCGCTCTCTTTTTAAGTGC 3' (*gmk*F_R2). The discriminatory power (D) based on the Simpson index of diversity and the 95 % confidence interval (CI) (95 % confidence the true D value lies within this range) were calculated for MLST data from all isolates (Grundmann *et al.*, 2001; Hunter & Gaston, 1988; Simpson, 1949). The D value and CI amongst isolates from each collection and amongst BSAC MRSA and MSSA isolates was also calculated.

2.4.2. Staphylococcal cassette chromosome *mec* (SCC*mec*) typing

Typing of all isolates was performed based on SCC*mec* sequences via the analysis of SCC*mec* PCR amplifications. A multiplex SCC*mec* PCR using 10 primer pairs was performed on all isolates under standard conditions (see materials and methods: section 2.3.1) with an annealing temperature of 53°C (Milheiriço *et al.*, 2007). Isolates belonging to SCC*mec* types I-V were detected by this PCR. PCR products (10 µl) were separated by electrophoresis on a 2 % agarose gel (see materials and methods: section 2.3.1) and the amplicon sizes were used to infer the SCC*mec* type as previously described (Milheiriço *et al.*, 2007).

2.4.2.1. SCC*mec* type IV subtyping

Isolates belonging to SCC*mec* type IV identified above were further typed into subtypes IVa to IVh (Milheiriço *et al.*, 2007). A multiplex PCR using seven primer pairs was performed on a subset of isolates under standard conditions (see materials and methods: section 2.3.1) with an annealing temperature of 48°C (Milheiriço *et al.*, 2007). PCR products (10 µl) were separated by electrophoresis on a 2 % agarose gel (see materials and methods: section 2.3.1) and the amplicon sizes were used to infer the SCC*mec* type IV subtype as previously described (Milheiriço *et al.*, 2007). Isolates that could not be assigned to SCC*mec* types I to V via the described protocol (Milheiriço *et al.*, 2007) were typed using one or more of the methods described in the sub-sections that follow. PCR products in each method were separated by electrophoresis as previously described (see materials and methods: section 2.3.1).

2.4.2.2. SCC*mec*-*orfX* junction observation

The junction between the SCC*mec* and *orfX* region was sequenced in a subset of isolates. PCR primers were designed by performing a multiple alignment of SCC*mec* type IV sequences obtained from the NCBI database (Appendix V: Group B) using MegAlign™ software version 8 (Lasergene®, DNASTAR, USA) using the clustalW method. A consensus sequence of approximately 1500 bp was generated and used to design PCR primers using Primer3 software. The resultant primer sequences were BLAST searched against the NCBI database and the primer pair with the highest similarity score with SCC*mec* type IV sequences (described above) was selected. The designed primers SCC*mec*-*orfX*F and SCC*mec*-

*orf*XR (MWG Eurofins) (Appendix IV) were utilised in a PCR performed under standard conditions. PCR products (15 µl) were separated by electrophoresis on a 2 % agarose gel (see materials and methods: section 2.3.1) and the amplicon sizes were used to infer the presence of the SCC*mec*.

PCR products were sequenced as described in the standard sequence determination protocol (see materials and methods: section 2.3.1). Two PCR primers were used initially and subsequently three nested sequencing primers were used to determine whether longer sequence reads could be obtained. These included one forward primer SCC*mec-orf*X2F and two reverse primers SCC*mec-orf*X2R and SCC*mec-orf*X3R (MWG Eurofins) (Appendix IV). Primers were designed using the SCC*mec-orf*X junction consensus sequence with the primer walking feature of SeqMan Pro™ software version 8 (Lasergene®). Contigs were assembled using SeqMan Pro™ software, aligned to the 1500 bp consensus sequence in MegAlign and BLAST searched against the NCBI database to determine the presence of the *orf*X region and SCC*mec* type..

2.4.2.3. *orf*X amplification

Amplification of the *orf*X region was performed for a subset of isolates. PCR primers were designed by performing a multiple alignment of *orf*X sequences associated with SCC*mec* type I-V obtained from the NCBI database (Appendix V: Group C) using BioEdit software and the clustalW method. A consensus sequence of approximately 480 bp was generated. The primers *orf*XF and *orf*XR (MWG Eurofins) (Appendix IV) were designed manually after no appropriate primers were obtained using Primer3 software. An annealing temperature of

50°C was used to amplify the PCR products. PCR products (10 µl) were separated by electrophoresis on a 2 % agarose gel (see materials and methods: section 2.3.1) and the amplicon sizes were used to infer the presence of the *orf*X.

2.4.2.4. SCC*mec* type VI PCR

Two PCRs were performed on a subset of isolates for the detection of isolates assigned to SCC*mec* type VI. Two primer pairs IS*1272*J-F with IS*1272*J-R and ccrB4-F with ccrB4-R were utilised from a previous study (Chen *et al.*, 2009) (Appendix IV). These primers amplify *mec* class B and *ccrB4* found in SCC*mec* type VI. An annealing temperature of 50°C was used to amplify the PCR products. PCR products (10 µl) were separated by electrophoresis on a 2 % agarose gel (see materials and methods: section 2.3.1) and the amplicon sizes were used to infer the presence of SCC*mec* type VI as previously described (Chen *et al.*, 2009).

2.4.2.5. SCC*mec* type VII PCR

A multiplex PCR using two primer pairs (*ccr*C2-F2 with *ccr*C2-R2 and *ccr*C8-F with *ccr*C8-R), targeting *ccr*C allotypes 2 and 8 respectively, was used on a subset of isolates. A second PCR using the primer pair *mec*C2-F with *mec*C2-R was utilised to target *mec* class C2 (Higuchi *et al.*, 2008) (Appendix IV). The annealing temperature for both PCRs was 50°C. PCR products (10 µl) were separated by electrophoresis on a 2 % agarose gel (see materials and methods: section 2.3.1) and the amplicon sizes were used to infer the presence of SCC*mec* type VII as previously described (Higuchi *et al.*, 2008).

2.4.2.6. SCC*mec* type VIII PCR

A subset of isolates belonging to SCC*mec* type VIII was identified by a previously described multiplex PCR (McClure *et al.*, 2010). The three primer pairs (*mecI*-F with *mecI*-R, *ccr*4-Fd with *ccr*4-R2 and VIII-F3 with VIII-R3) (Appendix IV) targeting the *mecI*, *ccr* complex type 4 and the recombination junction of SCC*mec* respectively were used. The multiplex PCR was performed using an annealing temperature of 52°C. PCR products (10 µl) were separated by electrophoresis on a 2 % agarose gel (see materials and methods: section 2.3.1) and the amplicon sizes were used to infer the presence of SCC*mec* type VIII as previously described (McClure *et al.*, 2010).

2.4.2.7. Investigation of SCC*mec* non-typeable isolates

Two multiplex PCRs were performed on isolates that were non-typeable by the above methods. Ten primer pairs were utilised from a previous study (Kondo *et al.*, 2007) (Appendix IV). The primers targeted the *mecA*, *ccrA1* to *ccrA4*, *ccrB1* to *ccrB4* and *ccr*C to enable assignment of the *ccr* complex type in each SCC*mec*. Four primer pairs in the second multiplex PCR targeted the *mecA*, *mecI*, IS*1272* and IS*431* to enable the assignment of the *mec* class. The multiplex PCRs were performed using an annealing temperature of 57°C and 60°C in the *ccr* type and *mec* class PCR respectively. PCR products (10 µl) were separated by electrophoresis on a 2 % agarose gel (see materials and methods: section 2.3.1) and the amplicon sizes were used to infer the presence of the above SCC*mec* features as previously described (Kondo *et al.*, 2007). The D value and the 95 % CI as previously proposed were calculated for SCC*mec* data from all isolates (Grundmann *et al.*, 2001; Hunter & Gaston,

1988; Simpson, 1949). The D value and CI were also calculated amongst isolates from each collection and amongst isolates belonging to each of the CCs: CC1, 5, 8, 22, 30 and 45.

2.4.3. *spa* typing

spa typing was performed on a subset of isolates, representative of 15 CCs, by determining the sequence of the polymorphic X region of protein A as described previously (Shopsin *et al.*, 1999). PCR was performed using an annealing temperature of 60°C and PCR products (5 µl) were separated by electrophoresis on a 2 % agarose gel and sequenced as described previously (see materials and methods: section 2.3.1). Chromatograms were assembled into contigs using BioNumerics software. The sequence of each amplicon was assigned to a *spa* type by comparison to the Ridom SpaServer database (http://spa.ridom.de) using BioNumerics software. Novel *spa* types were submitted to the Ridom SpaServer database. The *spa* type of each isolate was compared with MLST data and cluster analysis was performed on *spa* sequences using the Unweighted Pair Group Method with Arithmetic mean (UPGMA) method and BioNumerics software. The D and CI value were calculated for *spa* data from isolates belonging to CC1, 5, 8, 22, 30 and 45 (Grundmann *et al.*, 2001; Hunter & Gaston, 1988; Simpson, 1949).

2.5. Genetic diversity throughout the genome

2.5.1. Preliminary fluorescent amplified fragment length polymorphism (FAFLP)

FAFLP was performed empirically on a subset of isolates using four endonuclease combinations: *Eco*RI with *Mse*I, *Hin*dIII with *Hha*I, *Eco*RI with *Hha*I and *Bgl*II with *Csp*6I. An *E. coli K12* strain was used as a positive control in each FAFLP reaction. The

endonuclease digestion and adaptor ligation reactions for each combination are described below. Unless otherwise stated, the reagents for the digestion and ligation reactions were obtained from New England Biolabs, UK.

2.5.1.1. *Eco*RI with *Mse*I digestion

Genomic DNA (500 ng) was digested at 37°C for 3 h in a final reaction volume of 20.5 µl consisting of 10 U *Mse*I, 2 µl 10x NEBuffer 2, 0.2 µl bovine serum albumin (BSA, 10 g/mL) and 0.5 µl RNase A (30 mg/mL, Sigma-Aldrich). Five units of *Eco*RI (Fermentas, Thermo Scientific, UK) was subsequently added to each reaction mixture and incubated for a further 3 h at 37°C. Reactions were subsequently heated to 65°C for 10 min to inactivate the endonucleases and cooled to 4°C. Ligation was performed as described below (see ligation protocol: section 2.5.1).

2.5.1.2. *Hin*dIII with *Hha*I digestion

Genomic DNA (500 ng) was digested at 37°C for 2 h in a final reaction volume of 29 µl consisting of 6 U *Hin*dIII, 3 µl 10x NEBuffer 2 and 0.5 µl RNase A (30 mg/mL, Sigma-Aldrich). One µL master-mix containing six units of *Hha*I and 0.3 µl BSA (10 mg/mL) was subsequently added to each reaction mixture and incubated for a further 2 h at 37°C. Reactions were subsequently heated to 65°C for 10 min to inactivate the endonucleases and cooled to 4°C. Ligation was performed as described below (see ligation protocol: section 2.5.1).

2.5.1.3. *Eco*RI with *Hha*I digestion

Genomic DNA (500 ng) was digested at 37°C for 2 h in a final reaction volume of 22.5 µl consisting of 6 U *Hha*I, 3 µl 10x NEBuffer 2, 0.3 µl BSA (10 mg/mL) and 0.5 µl RNase A (30 mg/mL, Sigma-Aldrich). Six units of *Eco*RI was subsequently added to each reaction mixture and incubated for a further 2 h at 37°C. Reactions were subsequently heated to 65°C for 10 min to inactivate the endonucleases and cooled to 4°C. Ligation was performed as described below (see ligation protocol: section 2.5.1).

2.5.1.4. *Csp*6I with *Bgl*II digestion

Genomic DNA (500 ng) was digested at 37°C for 2 h in a final reaction volume of 15 µl consisting of 10 U *Csp*6I, 2 µl 10x Y+/Tango™ and 0.3 µl RNase A (30 mg/mL, Sigma-Aldrich). Three µL master-mix containing five units of *Bgl*II and 2.5 µl 10x Y+/Tango™ was subsequently added to each reaction mixture and incubated for a further 2 h at 37°C. Reactions were subsequently heated to 65°C for 10 min to inactivate the endonucleases and cooled to 4°C. Ligation was performed as described below (see ligation protocol: section 2.5.1).

2.5.1.5. Ligation protocol

Ligation was performed by adding 25 µl of master-mix containing 5 µl 10x T4 DNA-ligase buffer, 80 U T4 DNA ligase, 0.5 µl of each adaptor (MWG Eurofins) to the *Eco*RI with *Mse*I, *Hin*dIII with *Hha*I and *Eco*RI with *Hha*I double-digested DNA. Ligation was performed by

adding 27 µl of the above mastermix to the *Csp*6I with *Bgl*II digestion (Table 8). The reaction was incubated at 12°C for 16-18 h, heated to 65°C for 10 min to inactivate the ligase and cooled to 4°C. The ligation reactions were used as FAFLP PCR templates.

Table 8. Fluorescent Amplified Fragment Length Polymorphism (FAFLP) adaptors and sequences

Adaptor name	Adaptor sequence (5' to 3')	Concentration (µM)
EcoRI	CTCGTAGACTGCGTACC	2
	AATTGGTACGCAGTC	2
MseI	TACTCAGGACTCATC	20
	GACGATGAGTCCTGAG	20
*Hin*dIII	CTCGTAGACTGCGTACC	2
	AGCTGGTACGCAGTC	2
*Hha*I	GACGATGAGTCCTGATCG	20
	ATCAGGACTCATCG	20
*Bgl*II	CGGACTAGAGTACACTGTC	2
	GATCGACAGTGTACTCTAGTC	2
*Csp*6I	AATTCCAAGAGCTCTCCAGTAC	20
	TAGTACTGGAGAGCTCTTGG	20

The table displays double-stranded adaptor sequences and concentration utilised in FAFLP ligation reaction.

2.5.1.6. FAFLP PCR

A touchdown PCR (Desai *et al.*, 1998) was performed in 25 µl PCR reaction containing 2.5 µl of the ligated sample, 1x PCR reaction buffer, 2.5 mM $MgCl_2$, 0.2 mM each dNTP, 0.8 µM of each primer (Table 9) and 1.25 U *Taq* DNA polymerase recombinant. For each pair of primers, the forward primer was labelled at the 5'-end with one of the four fluorescent dyes 5'-carboxyfluorescein (FAM™),

2'-chloro-7'-phenyl-1, 4-dichloro-6-carboxyfluorescein (VIC®), 1',2'-benzo-4'-fluoro-7'-chloro-6-carboxy-4,7-dichlorofluorescein (NED™) and 4-(N, N-Dimethylaminoethylene) amino-N-allyl-1,8-naphthalimide (PET®) whilst the reverse primer was unlabelled (Life Technologies).

Table 9. FAFLP PCR primers and sequences

Primers	Sequence (5' to 3')
*Bgl*II+X*	GAGTACACTGTCGATCTX
*Csp*6I+X	GAGCTCTCCAGTACTACX
*Hin*dIII+X	GACTGCGTACCAGCTTX
*Hha*I+X	GATGAGTCCTGATCGCX
*Eco*RI+X	GACTGCGTACCAATTCX
*Mse*I+X	GATGAGTCCTGAGTAAX

***X represents a selective base: either nucleotide A, T, C or G.**

The primer sequences utilised on each double digest are described in Table 9. PCRs were performed with the primer pairs FAM-*Eco*RI+0 and *Mse*I+C (on *Eco*RI and *Mse*I digestions), FAM-*Hin*dIII+A and *Hha*I+T (on *Hin*dIII and *Hha*I digestions) and FAM-*Eco*RI+0 and *Hha*I+A (on *Eco*RI and *Hha*I digestions). PCRs using primers *Bgl*II+X (where X is nucleotide A, T, C or G labelled with FAM, PET, VIC and NED respectively) (Life technologies) and *Csp*6I+0, FAM-*Bgl*II+A and *Csp*6I+T, and FAM-*Bgl*II+0 and *Csp*6I+X were performed on *Csp*6I and *Bgl*II digestions. Unless otherwise stated the primers for FAFLP PCRs were obtained from MWG Eurofins.

The thermocycling conditions used were as follows:

Denaturation → 94°C → 2 min ⟩ 1 cycle

Denaturation → 94°C → 20 s

Annealing → 1 cycle → 66°C → 30 s

→ 9 cycles → decrease by 1°C increments →30 s → ⟩ 30 cycles

→ 20 cycles → 56°C → 30 s

Extension → 72°C → 2 min

Final extension → 60°C → 30 min ⟩ 1 cycle

FAFLP PCR reaction mixtures were cleaned up using the Agencourt® cleanSEQ® reagent as described previously (see materials and methods: section 2.3.1). One µL of cleaned product was added to 10 µl of Hi-Di formamide (Life technologies) and 0.5 µl of GeneScan™-600 LIZ® size standard (Life technologies). The size standard was labelled with a fifth proprietary fluorescent dye (LIZ). The reaction was heated at 95°C for 5 min and cooled to 4°C. Fragments were separated using the ABI 3730 Genetic Analyzer. POP-7™ was used for capillary electrophoresis which was run at 15 kV with a 15 s injection time and 1.6 kV injection voltage. A 48-capillary array with a length of 50 cm was utilised.

2.5.1.7. Preliminary FAFLP analysis

FAFLP AFs were separated on an ABI 3730 sequencer and the set of fragments generated for each sample were sized using GeneMapper software version 4.0 (Life Technologies). Based on the collective peaks from all samples, automatic bin assignment was performed. Process Quality Value tests were performed using the default peak quality settings. The size calling algorithm measures the similarity between the fragment pattern expected depending on the

size standard used and the actual peaks. Fragments ranging in size from 50 to 600 bp with a peak height of 50 relative fluorescent units (RFU) or above constituted the FAFLP profile of each isolate. Manual scoring of the fragments was then performed. Based on the collective AFs obtained from all samples, amplicon size categories (bins) greater than ±0.5 bp of one another were defined. Each bin represented an FAFLP locus and for each locus, the presence or absence of an AF was recorded in a binary format (1 and 0 for presence and absence respectively). The number of AFs in each profile and the number of common and polymorphic AFs between all profiles and those belonging to the same CC were analysed using the binary table. AFs were considered to be common to all isolates if they were present in ≥95 % of isolates. AFs that were present in ≥95 % of isolates of a specific lineage and in <5 % of the remaining isolates were classified as lineage-specific. AFs present in <95 % of all isolates or isolates of a specific CC were termed polymorphic.

To measure whether the distribution of fragments between 50 to 600 bp was skewed, the percentage of AFs less than and greater than 300 bp within each profile were calculated. Skewed profiles would increase the number of fragments of a similar size decreasing the accuracy of AF detection. The mean percentage of the genome represented by the AFs generated for each isolate was calculated. The mean *S. aureus* genome size was taken to be the mean genome size of 19 WGSs obtained from the NCBI database (Appendix V: Group D). Hierarchical cluster analysis was performed using statistiXL software version 1.8 (Roberts & Withers, 2009) using the UPGMA clustering method and the Jaccard co-efficient of similarity.

2.5.2. FAFLP

2.5.2.1. *Bgl*II+A with Csp6I+0 FAFLP analysis

Ligation reactions were performed for all isolates as detailed above using the endonucleases *Bgl*II and Csp6I. FAFLP PCR was performed as previously using the primers FAM-*Bgl*II+A and Csp6I+0 and data were analysed as described in the preliminary FAFLP analysis. FAFLP profiles constituted AFs between 50 to 600 bp and with a peak height of ≥50 RFU. The presence and absence of AFs among the FAFLP profiles were recorded in a binary format. Isolates that generated profiles with less than 45 or greater than 100 AFs were subjected to identification via determining the sequence of the 16S rRNA gene. Isolates that were identified as *S. aureus* were subjected to repeat FAFLP analysis from another DNA extract. A subset of isolates which exhibited >70 AFs were subjected to repeat FAFLP analysis. AFs constituting a profile from both replicates were compared manually and the percentage similarity was calculated by identifying the number of AF differences. The distribution of fragments between 50 to 600 bp and the mean percentage of the genome represented by each profile was calculated as previously.

Common and polymorphic AFs based on the definitions mentioned above were identified. *In silico* analysis of common AFs was performed using AFLP fragment predictor program, ALFIE (Underwood, 2008). *In silico* profiles were generated from 10 WGSs obtained from the NCBI database (Appendix V: Group E). A search was performed in ALFIE for fragments within ±3 bp of the expected AF size (21 bp less than that detected on the ABI 3730 sequencer due to the absence of the adaptor sequences) and present in each of the 10 WGSs. Sequences were obtained for the common AFs which were mapped to the WGSs

using the NCBI database. BLAST searches were performed of the AF sequences to identify features partially encoded in these regions. Multiple alignments of AF sequences from each of the 10 WGS were performed to calculate the percentage conservation of the sequence. This was calculated by observing the number of nucleotides that were conserved amongst the consensus sequence as compared with the number of nucleotides within the sequence. AFs present in ≥95 % of all MRSA or MSSA isolates and <5 % of the remaining isolates were identified as AFs exclusive to MRSA or MSSA. In addition, AFs present in ≥95 % of isolates from each collection were identified as common AFs whilst the remaining AFs were polymorphic. AFs present in ≥95 % of isolates from each collection and in <5 % of the remaining isolates were identified as exclusive to isolates from a specific collection.

2.5.2.2. Reproducibility studies

The reproducibility of the FAFLP assay was evaluated on a subset of isolates by performing FAFLP analysis in duplicate from different DNA extracts and from the same DNA extract.

2.5.2.3. Concordance of experimental FAFLP profiles with *in silico* profiles

In silico FAFLP profiles were generated for six strains with WGS data using ALFIE. The percentage concordance between *in silico* and experimental profiles for the six reference strains was determined.

2.5.2.4. Cluster analysis

Cluster analysis was performed on the 302 study isolates based on *Bgl*II+A with *Csp*6I+0 data. A dendrogram was generated using the UPGMA method based on the Dice similarity coefficient. Based on MLST data, a similarity cut-off was applied to the dendrogram to generate clusters. To display the FAFLP-based divergence, a further similarity cut-off was applied to generate sub-clusters. Profiles of isolates within one cluster were compared and the number of polymorphic AFs was calculated. These numbers related to the percentage divergence exhibited within a cluster. The number of polymorphic AFs between and within sub-clusters was also calculated. Clusters and sub-clusters were examined for correlations between FAFLP data and MLST or SCC*mec* typing. Isolates within a cluster were also studied for a correlation between FAFLP data and the source (collection) of the isolates. The D value and 95 % CI were calculated for FAFLP data, MLST and SCC*mec* data of all isolates (Grundmann *et al.*, 2001; Hunter & Gaston, 1988; Simpson, 1949).

2.5.2.5. FAFLP data correlation with lineage

Profiles of isolates assigned to one of seven CCs: CC1, 5, 8, 22, 30, 45 and 59 were compared. Common AFs within a CC were identified as fragments present in ≥95 % of isolates assigned to that CC. Conversely, polymorphic AFs within a CC were identified as fragments present in <95 % of isolates assigned to one CC. The FAFLP binary table was analysed using Stata software. Amongst isolates containing a particular AF, those AFs were identified where ≥95 % of the isolates belonged to a specific lineage. A manual search was subsequently performed for lineage-specific AFs; these fragments were present in >90 % of

isolates assigned to one CC and <5 % of isolates assigned to the remaining CCs. The sequences of lineage-specific AFs were determined experimentally or *in silico*.

2.5.2.6. Sequence determination of lineage-specific amplified fragments (AFs)

The identified lineage-specific AFs were sequenced. FAFLP PCRs were performed using the selective primers *Bgl*II+A with *Csp*6I+A on seven isolates (DBHT 98, 100, 198, 215, 230, 231, and 234) which were representative of seven CCs (CC1, 5, 8, 22, 30, 45 and 59). PCR products (15 µl) were separated by electrophoresis on a 4 % agarose gel at 105 V for 7 h. Each gel lane was sliced into five segments: 50-150 bp, 150-250 bp, 250-350 bp, 350-450 bp and 450-600 bp. PCR products were purified from the gel slices using the Wizard® SV gel and PCR clean-up system (Promega, UK). The weight of each gel slice was determined and 10 µl of membrane binding solution was added per 10 mg of gel slice. This solution was incubated at 80°C until the agarose dissolved. The solution was added to an SV minicolumn and incubated for 1 min. The solution was passed through the minicolumn, which contains a membrane that binds the DNA, by centrifugation at 16 000 x g for 1 min. The flow-through was discarded and 700 µl of membrane wash solution was added to the column and centrifuged as previously: this step was repeated with 500 µl of membrane wash solution and was centrifuged for 5 min. The flow-through was discarded and centrifuged for an additional 1 min. DNA was eluted by adding 5 µl of water to the membrane and centrifuging for 1 min.

Purified PCR products were utilised as templates in a further PCR using a BglII+A primer in combination with a two-base selective primer, *Csp*6I+AA, +AT, +AC or +AG. PCR products were separated by electrophoresis on a 4 % agarose gel at 105 V for 7 h. Fragments of

interest were excised from the gel using a pipette tip and transferred directly into sequencing reactions in a 96-well plate; sequencing was performed as mentioned previously (see materials and methods: section 2.3.1). Chromatograms were analysed using BioEdit software and sequences of lineage-specific AFs were BLAST searched against the NCBI database. A multiple alignment of search results was performed using BioEdit software and a consensus sequence was generated. PCR primers were designed from each consensus sequence using Primer3 software (Table 10).

PCR was performed for lineage-specific AFs in a 25 µl reaction under standard conditions (see materials and methods: section 2.3.1). The annealing temperature utilised in each PCR varied between 50 to 55°C (Table 10). PCRs were performed on 40 isolates representative of CC1, 5, 8, 15, 22, 30, 45, 59, 97 and 672 (Appendix VIII). PCR products were sequenced using the PCR primers following the sequence determination protocol described and resultant chromatograms were aligned using BioEdit software. PCRs designed for 11 lineage-specific AFs were validated on 48 isolates assigned to the 10 CCs listed above (Appendix VIII). These experiments were performed as blind assays i.e. the designated ST of the isolates were not known. Based on the lineage-specific sequence variations, isolates were assigned to a lineage.

Table 10. PCR primer sequences of lineage-specific amplified fragments (AFs)

Lineage-specific AF*	Specific to CC†	Primer name	Primer sequence (5' to 3')	Annealing temperature‡ (°C)
285 bp	CC5	285 bp_F	TCTGCTGATTGCTTTTGATACTAA	54
		285 bp_R	AAGATGATTCAAACGTAGAACGTG	
289 bp	CC8	289 bp_F	GAAGAAAGCTTTTATTCCTGTTT	50
		289 bp_R	TAACCATATCGGCTGCAGTG	
141 bp	CC22	141 bp_F	TTGTTTTGTTTGCCTCAACA	55
		141 bp_R	GCCTATAGGTATCAAAGTGACGAGA	
98 bp	CC30	98 bp_F	GGTTTTATCGTTTATAAAGGTGTTCG	54
		98 bp_R	TGCAATCTATGCTCTATTCCTGA	
156 bp	CC30	156 bp_F	TTCGTTATGCTTTGCGGTAA	50
		156 bp_R	GAAAGAGTAGTAACTTCAAAATCCGTA	
182 bp	CC30	182 bp_F	GGTACTGATAAGCCTGTTGATTG	52
		182 bp_R	AATGGCGTTATCCATTATGCT	
277 bp	CC30	277 bp_F	GCATTTGGCAAGTTAGCTTCT	50
		277 bp_R	CAAAAAGACGACCTACACAACA	
307 bp	CC30	307 bp_F	TGTGTTTAGGAACTGCATCTATCA	52
		307 bp_R	TAACCATATCGGCTGCAGTG	
389 bp	CC30	389 bp_F	GCCACAGCAAGACTTAGAT	50
		389 bp_R	CGTTTGTGTGGGGAATATG	
479 bp	CC30	479 bp_F	TCCAAATAATGCTGAACACTTAGC	52
		479 bp_R	TCATGCTACTTAAAATACAAGCACTG	
529 bp	CC30	529 bp_F	ACGCGCACGTATAGAGGTTT	54
		529 bp_R	CGGCACCTTTAGGTAGATGTTT	
82 bp	CC45	82 bp_F	TCGTACTATACCTCACTTTATTTACGG	55
		82 bp_R	AATCCAATAAGACAGAGTGAACG	
137 bp	CC45	137 bp_F	GCGTAGCGAAAGTATCTATGAAA	50
		137 bp_R	TCATGCAACTAGGTCAATTCG	
85 bp	CC59	85 bp_F	CGGAACGATACGATCATGTG	50
		85 bp_R	TTCATGAAGCTGTTGGTGAAG	
340 bp	CC59	340 bp_F	GTACAATGGGTTGCGGTCTT	50
		340 bp_R	TCAGTTATAAGAAACGTTTTGCTG	

*Lineage-specific amplified fragments (AFs) were identified from FAFLP reactions using the primers *BglI*1+A with *Csp*6I+0.

†Lineages (CCs) to which AFs of a precise size were specific.

‡Annealing temperature for each PCR performed under standard conditions (see materials and methods: section 2.3.1).

2.5.2.7. *In silico* determination of lineage-specific AFs

The sequence of the lineage-specific AFs that were not obtained experimentally were determined *in silico*. The genomic loci of the AFs were determined using *in silico* profiles generated from WGSs and ALFIE. Fragment sizes in ALFIE were 21 bp less than those determined on the ABI 3730 Genetic Analyzer due to the absence of the adaptor sequences. Fragments that were ±7 bp of the expected size in ALFIE were identified and the fragment closest in size to the expected was selected. Based on the locus of the fragment, the sequence was obtained from WGSs in the NCBI database. BLAST search results of these sequences from WGSs representative of 10 lineages (CC1, 5, 8, 22, 30, 59, 93, 151, 398 and 425) were aligned using BioEdit software.

2.5.2.8. Lineage-specific AF sequence variation

Lineage-specific AF sequence alignments were analysed to identify lineage-specific SNPs (substitution), point mutations (1 bp insertion and deletion) and indels (>1 bp). Alignments generated from experimental sequences were analysed for the presence of specific variations within the following CCs: CC1, 5, 8, 15, 22, 30, 45, 59, 97 and 672. Alternatively, alignments from WGSs were analysed for the presence of specific variations within CC1, 5, 8, 22, 30, 59, 93, 151, 398 and 425. The number of SNPs, point mutations and indels were compared between each of the AF sequences. The numbers of sequence variations specific to isolates of one CC were identified.

2.5.2.9. Proteins encoded on lineage-specific AF sequence and their function

BLAST search results were performed of the consensus sequences of lineage-specific AFs. Based on annotated WGSs, partial open-reading frames (ORFs) encoded in these consensus sequences were determined. The proteins encoded by these ORFs and their possible functions were researched in the literature. In addition, the translated amino acid sequence of these ORFs from WGSs belonging to CC1, 5, 8, 22, 30, 59, 93, 151, 398 and 425 were aligned against the lineage-specific AF sequence alignments. The SNPs or point mutations found in ORFs were compared against the amino acid sequences to determine whether the variations were synonymous (no change in amino acid sequence) or non-synonymous (change in amino acid sequence).

2.5.2.10. Pyrosequencing™

A search was performed amongst AFs displaying lineage-specific variations as a target for a Pyrosequencing™ assay. AFs displaying lineage-specific variations for the maximum number of lineages were selected. The sequences of these AFs were compared to identify an approximately 100 bp region with variations specific to multiple lineages. The region amongst the AFs with the highest number of lineage-specific variations was selected for Pyrosequencing™. PCR primers were designed manually using BioEdit software to amplify this region. A hot start PCR was performed in a 25 µl PCR reaction containing 1x HotStarTaq™ master mix (Qiagen, UK), 0.12 µM of unlabelled primer, 0.14 µM of biotinylated primer (MWG Eurofins), 0.012 M of betaine and 75 ng of DNA. The forward primer sequence was labelled at the 5' end with biotin. The forward primer, 307 bp_pyro_F,

sequence was 5' TGTGTTTAGGAACTGCATCTATCA 3' and the reverse primer, 307 bp_pyro_R, sequence was 5' ACTAAGTATTCCAAAAATGCATGA 3'. Thermocycling was performed on GeneAmp® PCR system 9700 (Life Technologies) and cycling conditions used were as follows: 95°C for 15 min; 95°C for 2 min; followed by 35 cycles of 95°C for 45 s, 55°C for 45 s and 72°C for 1 min and a final extension at 72°C for 5 min with a hold cycle at 4°C. PCR products (2 µl) were confirmed on a 2 % agarose gel as previously described.

A 60 µl solution was made up of 20 µl of PCR reaction, 3 µl of Streptavidin Sepharose™ High Performance beads (GE Healthcare Life Sciences, UK), 40 µl of PyroMark Binding Buffer and 17 µl of water. All reagents were obtained from Qiagen, UK unless otherwise stated. Each reaction was incubated for 10 min at room temperature whilst shaking at 140 x g. The solution was aspirated using the PyroMark™ Vacuum Prep Workstation, resulting in the beads adhering to the vacuum tool. The tool was then transferred to a solution of 70 % ethanol which was aspirated for 5 s. This process was repeated for solutions of PyroMark Denaturation Solution and 1x PyroMark Wash Buffer sequentially. The tool was held vertically to ensure all liquid was drained from the tubing. Once the vacuum was released, the tool was placed in a PSQ 96 Thermoplate Low containing 45 µl of sequencing primer (0.3 µM) in PyroMark Annealing buffer. Two forward sequencing primers were used on PCR products from each isolate. The primer sequences were 5' CATAAAACGACAATTTCA 3' for 307 bp_pyro_F1 and 5' CATAAAACAACAATTTCA 3' for 307 bp_pyro_F2. The PSQ plate was incubated at 80°C for 2 min and allowed to cool slowly at room temperature. The vacuum tool was washed thoroughly by aspirating distilled water for 30 s.

Pyrosequencing™ was performed using PyroMark Gold Q96 reagents on the automated PyroMark Q96 ID system. One vial each of enzyme and substrate were reconstituted in 620 µl of distilled water. The PyroMark Q96 ID system was utilised with the PyroMark ID software version 1.0 (Biotage, UK). An SQA run was selected and the nucleotide dispensation order was programmed as 25 cycles of T, C, G and A. The instrument parameters were set to a nucleotide dispensation pressure of 650 mbar, reagent dispensation of 400 mbar and a 65 s cycle time. The volume of enzyme, substrate and the four dNTPs needed for the run was calculated by the software and added to the PSQ 96 reagent cartridge. Reagents were automatically dispensed by the instrument according to the nucleotide dispensation order. Sequences were displayed as a Pyrogram™ and in text format. Sequences were aligned against the expected lineage-specific AF sequence using BioEdit software. Based on previously identified lineage-specific variations, isolates were assigned to a lineage.

2.5.2.11. FAFLP comparison with *spa* typing and whole-genome mapping (WGM)

FAFLP data were compared with *spa* typing and WGM data for a subset of isolates. The percentage of the genome sampled by each method was compared as were the number of profiles or *spa* types exhibited with each method. The percentage similarity displayed amongst isolates assigned to the same lineage was calculated for each method based on cluster analysis using the UPGMA method. Optical maps generated a restriction fragment patterns across the genome, portions of these patterns were identified that were unique to isolates of one lineage. The number of lineage-specific AFs and restriction fragment patterns identified using FAFLP and WGM respectively were compared. The proportion of the

genome that was represented by lineage-specific regions identified using FAFLP and WGM were used to calculate the percentage of the genome specific to one lineage.

2.5.3. Insertion sequence targeted FAFLP (IS-FAFLP)

An IS targeted FAFLP (IS-FAFLP) assay was developed for *S. aureus*. A search of WGSs for an IS was performed. To ensure that the sequence is relatively conserved between lineages, the parameters of the target IS were that the sequence should be present in all *S. aureus* genomes but belong to one IS family. In addition, the IS should be present at variable loci and in a variable copy number amongst the WGSs. Firstly, the presence of ISs were compared amongst 15 WGSs from the NCBI database (Appendix V: Group F). The search terms 'insertion' and 'sequence' or 'transposase' were used to identify potential ISs against the genome annotations. Following this, a multiple alignment was performed on the 15 WGSs using Mauve software version 2.3.1. Using the find features tool in Mauve software, all genomes were searched for features containing the keywords 'insertion' and 'sequence' or transposase'. The search results of these terms from Mauve software were BLAST searched against the NCBI database. Search results from Mauve software that displayed similar sequences in the most number of other WGSs were identified as potential FAFLP targets. The BLAST search results of potential FAFLP target sequences were aligned using Bioedit software and a consensus sequence was generated.

Restriction endonuclease recognition sequences within the target consensus sequences were determined *in silico* using the Enzymes-Restriction Map method in GeneQuest™. The search was limited to those restriction endonucleases that recognise a 4 bp sequence. This method identified restriction endonuclease recognition sequences for up to two endonucleases per FAFLP target. Reverse primers were designed manually using BioEdit software; primers were

designed within the target sequence upstream of the restriction endonuclease site. An IS-targeted FAFLP was developed and performed on a subset of isolates. The assay was performed with minor modifications to the previously described FAFLP method. Genomic DNA was digested with a frequent-cutting endonuclease. FAFLP PCR was performed using the transposase-specific reverse primers on *Eco*RI with *Mse*I and *Hin*dIII with *Hha*I ligation reactions performed as described previously. PCR products (10 µl) were separated by electrophoresis on a 2 % agarose gel (see materials and methods: section 2.3.1). FAFLP PCR reactions performed on *Hin*dIII with *Hha*I ligation reactions were separated on an ABI 3730 sequencer and analysed using GeneMapper software.

2.5.4. WGM

WGM was performed on a subset of isolates by OpGen® Inc. (Gaithersburg, USA). In brief, bacterial cells were gently lysed using a protocol developed by OpGen® to obtain high molecular weight DNA fragments (250 kb or more in size). The DNA molecules were restricted with a rare-cutting endonuclease, *Noc*I which recognises the following sequence:

5' C^CATGG 3'

3' GGTAC^C 5'

As the DNA was restricted, gaps were formed between the fragments as the molecules were stretched taut. The fragments were fluorescently labelled and assigned a size based on the fluorescent intensity of each fragment. The order and number of fragments and the size of each fragment formed a single molecule map. The single molecule restriction maps were aligned to produce a consensus map covering the entire genome (optical map). Analysis of the generated optical maps was performed using MapSolver™ software version 3.1. The

maps were visualised in the software which displayed the number of restriction fragments, their order and location in the genome. Maps were aligned with *in silico* maps from 21 WGSs obtained from the NCBI database (Appendix V: Group G). Multiple alignments were performed for all isolates of a specific CC, one representative from each CC and amongst single isolates of ST8, ST239 and ST30. Fragments unique to one map in the alignment were highlighted as were fragments common amongst the optical maps. The MapSolver™ software algorithm also highlighted fragments between optical maps, which may indicate an inversion event. Cluster analysis was performed on optical maps and *in silico* maps. Comparative genomics was performed to identify fragment patterns unique to each isolate and between isolates of a single CC. Comparison of lineage-specific fragment patterns was performed against *in silico* maps of strains belonging to the same CC to identify possible genes and features encoded in these regions.

CHAPTER 3

3. Results

3.1. *S. aureus* identification

3.1.1. 16S rRNA gene sequence determination

Eight isolates were identified to the species level based on the16S rRNA gene sequence. The PCR primers amplified approximately 1500 bp of the 16S rRNA gene and each sequencing read ranged between 700 to 800 bp. The quality of the reads was determined by the QV20+ value (number of bases which have a base caller quality value of ≥20), as determined by ABI Sequence Scanner software version 1.0 (Life Technologies). Forward and reverse sequencing reads were aligned into contigs which were sized between 600 to 700 bp. The eight isolates (DBHT 1 to 8) from SRL required confirmatory identification as DBHT 4 and 5 were negative for the SCC*mec* element whilst the other six isolates from the same patient were positive. All isolates were confirmed as *S. aureus*. All positive controls were appropriately identified as *E. coli* whilst no amplicons were detected amongst negative controls.

3.1.2. Matrix-assisted laser desorption ionization-time-of-flight mass spectrometry (MALDI-TOF)

Spectra were obtained for all 32 isolates analysed with the MALDI mass spectrometer. Analysis of the spectra using MALDI Biotyper software identified all isolates as *S. aureus*. The top ten matches to the database for each spectrum were *S. aureus*; matches produced a score of 2.3 or above. All positive controls were appropriately identified as *E. coli* whilst no mass peaks were detected amongst negative controls. All 32 spectra were analysed for the presence and absence of 13 m/z peaks, detailed by Wolters *et al.* (2011), for assignment to one of 15 MALDI types. Each type corresponded to one of the CCs: CC5, 8, 22, 30 or 45. MALDI type assignment corresponded to MLST assignment for 30 of the 32 isolates examined. One

of the remaining isolates was assigned to MALDI type m002 which is indicative of ST1, 15 or 80. However, MLST data assigned this isolate to ST59. The remaining isolate produced a MALDI type which was not described in the Wolters *et al.* (2011) study: MLST data assigned this isolate to CC672 (Table 11).

3.2. Characterisation of *S. aureus*

3.2.1. Multilocus sequence typing (MLST)

Multilocus sequence typing was performed on all 302 isolates by determining the sequence of a region within seven housekeeping genes which ranged in size from 402 to 516 bp (Enright *et al.*, 2000). Closely related STs are displayed as clusters known as CCs that aggregate around the 'founder' type which is the most closely related ST to all of the STs within the cluster. Each isolate was assigned to an allelic profile, ST and CC. The 302 isolates belonged to one of 62 STs which were assigned to 20 CCs and eight singletons (Figure 20). Up to eight different STs were identified amongst CC1, 5, 8, 22, 30 and 45. The remaining CCs were represented by up to three STs. Of two hundred and thirty-nine isolates in this study, 128 were CC22 (42 %), 36 were CC30 (12 %), 30 were CC8 (10 %), 23 were CC5 (8 %) and 22 were CC45 (7 %). The remaining isolates belonged to 15 CCs and eight singletons. Twenty-two novel STs were identified amongst the isolates assigned to CC1, 5, 8, 22, 30, 45 and 59. ST2042 was conferred by a novel combination of seven previously described alleles whilst ST2221 was conferred by a profile with four novel alleles out of seven. The remaining 20 novel STs displayed a unique allele at one of the seven loci (Table 12). The *gmk* primers that were redesigned for this study produced contigs spanning the entire target region and enabled the submission of three novel STs with a novel *gmk* allele.

Table 11. Clonal complex assignment of isolates based on matrix-assisted laser desorption ionization-time-of-flight mass spectrometry types

Isolate	m/z values*													MALDI type†	CC/ST based on MALDI type	MLST‡	
	3276	3876	4511	4641	5002	5032	5419	5437	5508	5524	6591	6612	7734			ST	CC
DBHT 1	0	1	0	0	0	1	0	1	0	1	0	1	0	NT§	NT	361	CC672
DBHT 4	1	1	0	0	1	0	0	1	0	1	0	1	0	m008	CC22	22	CC22
DBHT 5	1	1	0	0	1	0	0	1	0	1	0	1	0	m008	CC22	22	CC22
DBHT 7	1	1	0	0	1	0	0	1	0	1	0	1	0	m008	CC22	22	CC22
DBHT 10	1	1	1	0	0	1	1	0	1	0	1	1	0	m014	CC30	30	CC30
DBHT 15	0	1	0	0	0	1	0	1	0	1	1	1	0	m009	CC8	8	CC8
DBHT 16	0	1	0	0	0	1	0	1	0	1	1	1	0	m009	CC8	8	CC8
DBHT 17	1	0	0	0	0	1	0	1	0	1	0	1	0	m001	CC5	5	CC5
DBHT 18	1	1	0	0	1	0	0	1	0	1	0	1	0	m008	CC22	22	CC22
DBHT 19	1	1	0	0	0	1	0	1	0	1	0	1	0	m002	ST1, 15 or 80	1	CC1
DBHT 24	1	1	0	0	1	0	0	1	0	1	0	1	0	m008	CC22	22	CC22
DBHT 26	0	1	0	0	0	1	0	1	0	1	1	1	0	m009	CC8	250	CC8
DBHT 32	1	1	0	0	1	0	0	1	0	1	0	1	0	m008	CC22	22	CC22
DBHT 36	1	1	1	0	0	1	1	0	1	0	0	1	0	m013	CC30	39	CC30
DBHT 39	1	1	1	0	0	1	0	1	0	1	0	0	1	m003	CC45	2032	CC45
DBHT 43	1	0	0	0	0	1	0	1	0	1	0	1	0	m001	CC5	5	CC5
DBHT 45	1	1	0	0	1	0	0	1	0	1	0	1	0	m008	CC22	22	CC22
DBHT 51	1	1	1	0	0	1	1	0	1	0	0	1	0	m013	CC30	36	CC30
DBHT 54	1	0	0	0	0	1	0	1	0	1	0	1	0	m001	CC5	5	CC5
DBHT 59	1	0	0	0	0	1	0	1	0	1	0	1	0	m001	CC5	5	CC5
DBHT 65	0	1	0	0	0	1	0	1	0	1	1	1	0	m009	CC8	630	CC8
DBHT 73	1	1	0	0	0	1	0	1	0	1	0	1	0	m002	ST1, 15 or 80	15	CC15
DBHT 76	0	1	0	0	0	1	0	1	0	1	1	1	0	m009	CC8	239	CC8
DBHT 84	1	1	0	0	0	1	0	1	0	1	0	1	0	m002	ST1, 15 or 80	1	CC1
DBHT 87	1	1	1	0	0	1	0	1	0	1	0	0	1	m003	CC45	45	CC45
DBHT 119	1	1	0	0	1	0	0	1	0	1	0	1	0	m008	CC22	22	CC22

Table 11 continued. Clonal complex assignment of isolates based on matrix-assisted laser desorption ionization-time-of-flight mass spectrometry types

Isolate	m/z values*													MALDI type†	CC/ST based on MALDI type	MLST‡	
	3276	3876	4511	4641	5002	5032	5419	5437	5508	5524	6591	6612	7734			ST	CC
DBHT 151	1	1	0	0	1	0	0	1	0	1	0	1	0	m008	CC22	22	CC22
DBHT 204	1	1	0	0	0	1	0	1	0	1	1	1	0	m010	CC8	8	CC8
DBHT 230	1	1	0	0	0	1	0	1	0	1	0	1	0	m002	ST1, 15 or 80	59	CC59
DBHT 236	1	1	1	0	0	1	1	0	1	0	0	1	0	m013	CC30	30	CC30
DBHT 242	1	1	1	0	0	1	1	0	1	0	0	1	0	m013	CC30	30	CC30
DBHT 266	1	1	0	0	0	1	0	1	0	1	1	1	0	m010	CC8	72	CC8

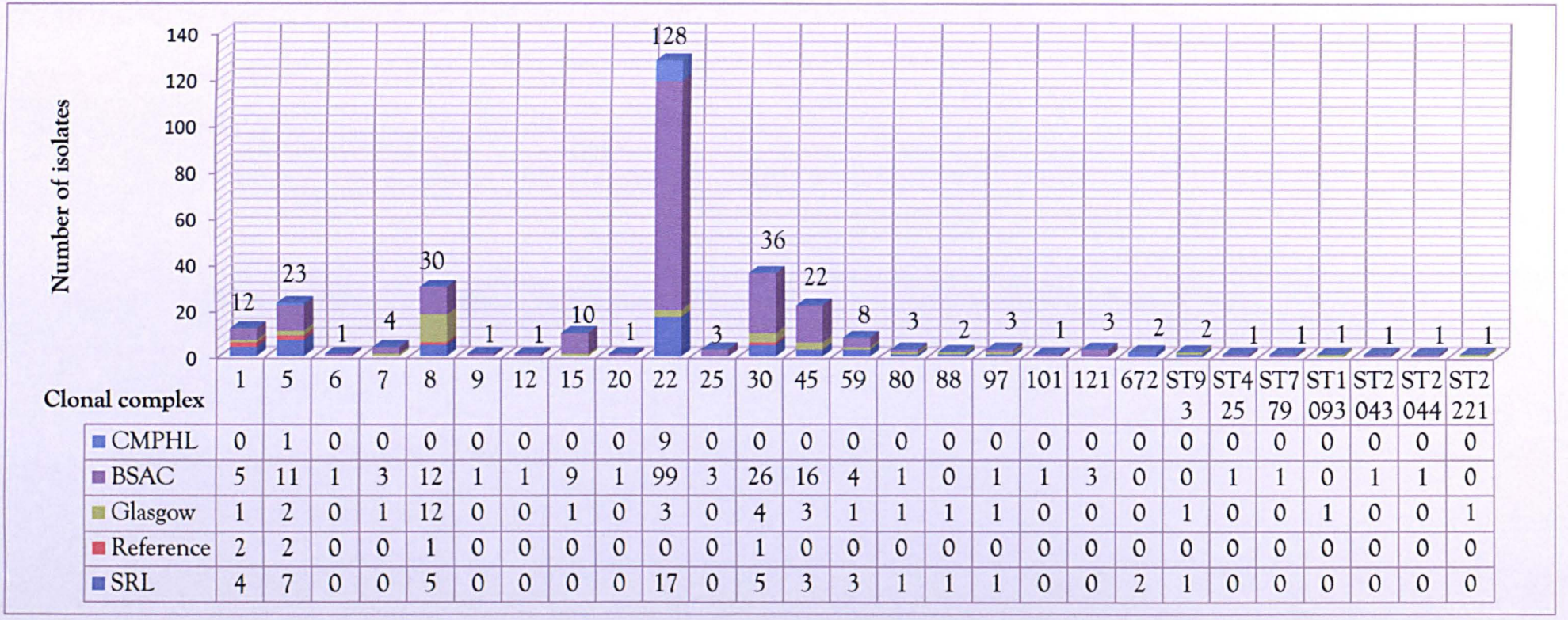
*m/z values, mass to charge ratio.

†Presence or absence of 13 m/z peaks for assignment of MALDI type (Wolters *et al.*, 2011).

‡MLST-based assignment to sequence type (ST) and clonal complex (CC).

§NT, non-typeable.

Figure 20. MLST based clonal complex assignment



Clonal complex	1	5	6	7	8	9	12	15	20	22	25	30	45	59	80	88	97	101	121	672	ST93	ST425	ST779	ST1093	ST2043	ST2044	ST2221
CMPHL	0	1	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BSAC	5	11	1	3	12	1	1	9	1	99	3	26	16	4	1	0	1	1	3	0	0	1	1	0	1	1	0
Glasgow	1	2	0	1	12	0	0	1	0	3	0	4	3	1	1	1	1	0	0	0	1	0	0	1	0	0	1
Reference	2	2	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SRL	4	7	0	0	5	0	0	0	0	17	0	5	3	3	1	1	1	0	0	2	1	0	0	0	0	0	0

Displays clonal complex assignment, based on MLST data, amongst isolates from each of five collections.

*S, singleton sequence type.

Table 12. Novel multilocus sequence types identified amongst isolates in this study

Isolate	*arcC*[*]	*aroE*	*glpF*	*gmk*	*pta*	*tpi*	*yqiL*	ST[†] assignment	CC[‡]	SCC*mec* type
DBHT 196	10	14	8	6	10	230	2	2025	CC45	N/A[§]
DBHT 211	1	4	1	4	12	231	10	2026	CC5	N/A
DBHT 13	22	1	14	23	224	4	31	2027	CC88	IVa
DBHT 37	3	295	1	1	4	4	3	2028	CC8	N/A
DBHT 53	5	4	1	4	4	6	229	2029	CC7	IVb/F
DBHT 56	1	3	1	14	11	51	230	2030	CC80	IVc
DBHT 141	7	6	1	5	8	229	6	2031	CC22	IVh
DBHT 39	220	14	8	6	10	3	2	2032	CC45	IVg
DBHT 41	220	14	8	6	10	3	2	2032	CC45	IVg
DBHT 232	6	298	6	2	7	17	19	2033	CC121	N/A
DBHT 275	2	297	2	2	6	3	2	2034	CC30	N/A
DBHT 261	10	14	271	6	10	3	2	2035	CC45	N/A
DBHT 200	1	4	1	4	225	1	10	2036	CC5	N/A
DBHT 168	7	296	1	5	8	8	6	2037	CC22	IVh
DBHT 21	7	6	1	5	8	228	6	2038	CC22	IVh
DBHT 9	1	1	1	159	1	1	1	2039	CC1	N/A
DBHT 248	1	4	1	161	12	1	10	2040	CC5	N/A
DBHT 271	19	23	15	2	19	232	15	2041	CC59	N/A
DBHT 194	8	2	2	27	6	3	2	2042	CC30	N/A
DBHT 221	36	14	270	150	107	116	105	2043	Singleton	N/A
DBHT 228	36	298	43	150	107	116	105	2044	Singleton	N/A
DBHT 38	238	2	2	2	6	3	2	2220	CC30	N/A
DBHT 63	23	321	269	161	1	47	231	2221	Singleton	N/A

Novel multilocus sequence types, consisting of novel alleles (highlighted in pink) at one or more loci or a unique combination of previously described alleles, were submitted to the MLST database.

[*]Numbers represent allele numbers at one of seven loci (***arcC*, *aroE*, *glpF*, *gmk*, *pta*, *tpi*** or ***yqiL*).**

[†]ST, sequence type.

[‡]CC, clonal complex.

[§]N/A indicates isolates were MSSA whilst the remaining isolates were MRSA.

The 50 isolates amongst the SRL collection belonged to 20 STs and were assigned to 11 CCs and one singleton; six of the STs were novel. Ninety percent of isolates belonged to one of the eight major lineages. The distribution of isolates to eight lineages was as follows: 17 were CC22 (34 %), seven were CC5 (14 %), five were CC8 (10 %), five were CC30 (10 %), four were CC1 (8 %), three were CC45 (6 %), three were CC59 (6 %) and one was CC80 (2 %) (Figure 20). The remaining 10 % of isolates belonged to three CCs and one singleton. Of the fifteen isolates assigned to EMRSA-15 by SRL, 13 belonged to CC22 and the remaining two were assigned to CC672 (Table 13). Thirty-five of the remaining 50 isolates were non-typeable, displayed a specific pattern or displayed a distinct pattern. Those isolates where no data were available by bacteriophage typing and/or PFGE belonged to one of 10 CCs and a singleton ST.

The Reference collection was assigned to ST1, 5, 36 or 250 and these results were concordant with the STs previously described from WGSs (Saunders *et al.*, 2004). The Glasgow collection was assigned to 20 STs corresponding to 12 CCs and three singletons; three of the STs were novel. For this collection of isolates, 22 isolate (65 %) were assigned to one of four CCs: 12 were CC8 (35 %), four were CC30 (12 %), three were CC22 (9 %) and three were CC45 (9 %). The remaining isolates belonged to eight CCs and three singleton STs. Based on previous PFGE analysis and MLST, five non-typeable isolates belonged to three CCs and one singleton, two isolates exhibiting the same PFGE type belonged to CC22 and 30, and 21 isolates exhibiting distinct PFGE types were assigned to 10 CCs and one singleton. Of the 34 isolates, PFGE assigned three to EMRSA-15 (belonging to CC22) and four to EMRSA-16 (belonging to CC30). The remaining 27 isolates belonged to 11 CCs and three singletons.

Table 13. Multilocus sequence typing (MLST) data concordance with Staphylococcal Reference Laboratory collection

Isolate number	Bacteriophage typing results	PFGE result	MLST result	Comments
DBHT 1	NT*	EMRSA-15	CC672	From single patient over six months with wound abscesses and recurrent bacteraemia
DBHT 2	NT	EMRSA-15	CC672	
DBHT 3	NT	EMRSA-15	CC22	
DBHT 4	NT	Pattern A	CC22	
DBHT 5	ND†	ND	CC22	
DBHT 6	NT	EMRSA-15	CC22	
DBHT 7	EMRSA-15	EMRSA-15	CC22	
DBHT 8	EMRSA-15	EMRSA-15	CC22	
DBHT 9	Distinct strain	Distinct strain	CC1	
DBHT 12	NT	Pattern A	CC80	
DBHT 13	Distinct strain	Distinct strain	CC88	
DBHT 20	ND	EMRSA-15	CC22	
DBHT 21	ND	EMRSA-15	CC22	
DBHT 22	ND	EMRSA-15	CC22	
DBHT 23	ND	EMRSA-15	CC22	
DBHT 24	ND	EMRSA-15	CC22	
DBHT 31	EMRSA-15	EMRSA-15	CC22	From same hospital
DBHT 32	EMRSA-15	EMRSA-15	CC22	
DBHT 33	NT	EMRSA-15	CC22	From same hospital
DBHT 34	NT	EMRSA-15	CC22	
DBHT 35	Distinct strain	ND	CC30	From the same hospital ward
DBHT 36	Distinct strain	ND	CC30	
DBHT 37	Distinct strain	ND	CC8	
DBHT 38	Distinct strain	ND	CC30	
DBHT 39	Distinct strain	Pattern C	CC45	From same hospital
DBHT 40	Distinct strain	Pattern A	CC5	
DBHT 41	Wide-spread sporadic strain	Pattern B	CC45	
DBHT 42	Distinct strain	Distinct strain	CC5	From same hospital
DBHT 43	Distinct strain	Distinct strain	CC5	
DBHT 44	Pattern A	ND	CC5	From same hospital
DBHT 45	Pattern A	ND	CC20	
DBHT 46	Pattern A	Pattern A	CC1	
DBHT 47	NT	ND	CC97	
DBHT 48	Distinct strain	Distinct strain	CC8	
DBHT 49	Pattern A	ND	CC59	From same hospital
DBHT 50	Pattern A	ND	CC59	

Clonal complex (CC), based on MLST data, concordance with bacteriophage typing and PFGE typing results for the SRL collection which included epidemic MRSA (EMRSA).

Distinct strains displayed bacteriophage or PFGE profiles which were different from previously obtained profiles.

*NT, non-typeable.

†ND, no data.

The BSAC collection was assigned to 44 STs which belonged to 18 CCs and four were singletons (Figure 20). The 99 MSSA isolates were assigned to 37 STs of which 11 were novel SLVs of ST5, 30 or 45. These 37 STs were assigned to 17 CCs and three singletons. The 103 MRSA isolates were assigned to 10 STs and two of these were novel SLVs of ST22. These 10 STs were assigned to CC5, 22, 30, 45 or 80 and one was a singleton (Table 14). Of the 96 and six isolates assigned to ST22 (48 %) and ST36 (3 %) respectively, the majority (≥93 %) were MRSA.

Ten isolates, including eight from an ICU outbreak, were obtained from CMPHL. Nine isolates (90 %), including six from ICU, were assigned to ST22 (Figure 20). Five isolates were identified as EMRSA-15 by bacteriophage or PFGE typing and belonged to ST22.

The D value for MLST data was 0.817 from the 302 isolates with a CI of 0.773 to 0.860. The D values and CI values were calculated for individual isolate collections. The CI values overlapped between three collections; no overlap between the Glasgow and BSAC collection or the CMPHL and the remaining four collections was observed (Table 15). The D value for the BSAC MRSA and MSSA isolates was 0.257 and 0.990 with a CI of 0.139 to 0.364 and 0.972 to 1.000 respectively. The CI values between the BSAC MRSA and MSSA did not overlap.

Table 14. Distribution of meticillin-resistant *S. aureus* (MRSA) and meticillin-susceptible *S. aureus* (MSSA) British Society for Antimicrobial Chemotherapy isolates within clonal complexes

CC	Prevalence in CC (%)	MRSA prevalence (%)	MSSA prevalence (%)
CC1	2	0	5
CC5	5	1	10
CC6	1	0	1
CC7	1	0	3
CC8	6	0	12
CC9	1	0	1
CC12	1	0	1
CC15	4	0	9
	1	0	1
CC22	48	89	8
CC25	1	0	3
CC30	13	7	19
CC45	8	1	15
CC59	2	0	4
CC80	1	1	0
CC97	1	0	1
CC101	1	0	1
CC121	1	0	3
Singleton	2	1	3
Total (%)	100	100	100

The table displays prevalence of all *S. aureus*, MRSA and MSSA assigned to each clonal complex (CC) in the BSAC collection.

Table 15. Discriminatory power and confidence intervals based on MLST data for five collections

Isolate collection	D value	CI value
SRL	0.878	0.806-0.950
Reference	0.867	0.738-0.995
Glasgow	0.948	0.911-0.985
BSAC	0.761	0.699-0.824
CMPHL	0.2	0.000-0.504

The discriminatory power (D) (Simpson, 1949) and confidence interval (CI) (Grundmann *et al.*, 2001) values were calculated based on MLST data amongst isolates obtained from each collection.

3.2.2. Staphylococcal cassette chromosome *mec* (SCC*mec*) typing

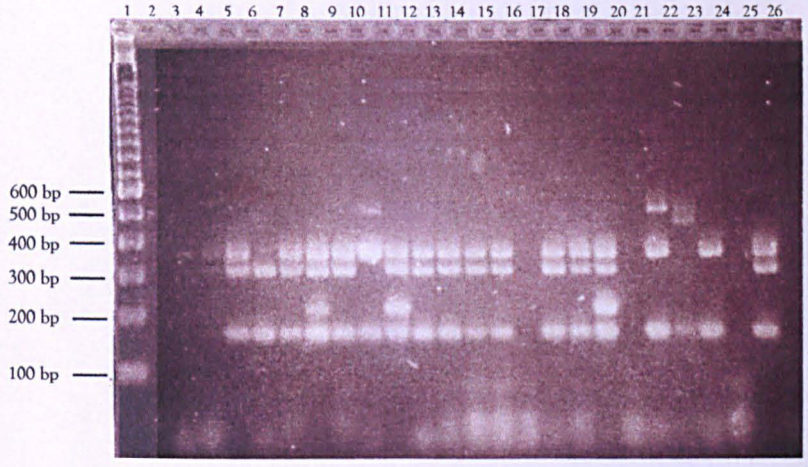
3.2.2.1. SCC*mec* typing I to V

Based on the presence of complete or partial SCC*mec* sequences, 184 of 302 *S. aureus* isolates (60.9 %) were identified as MRSA. The majority of MRSA (n=174, 94.6 %) isolates belonged to SCC*mec* types I to V (Figure 21); 128 of these MRSA (73.6 %) were type IV, 23 (13.2 %) were type II, 13 (7.5 %) were type V, six (3.4 %) were type I and four (2.3 %) were type III.

Of the eight isolates obtained from the same patient, six belonged to SCC*mec* type IV and no element was detected in the remaining two isolates. To confirm the absence of SCC*mec*, these two isolates along with one other MRSA isolate from the same patient were subjected to 10 individual PCRs based on the primers pairs from the multiplex. The MRSA isolate was utilised as a positive control. Each of the 10 PCRs yielded no product from either isolate.

The positive control produced PCR products in the PCRs targeting the *mecA* and *ccrB2*; these corresponded to the presence of SCC*mec* type IV and concurred with the result obtained with the multiplex PCR.

Figure 21. SCC*mec* type I-V determination based on size variations in SCC*mec* amplicon



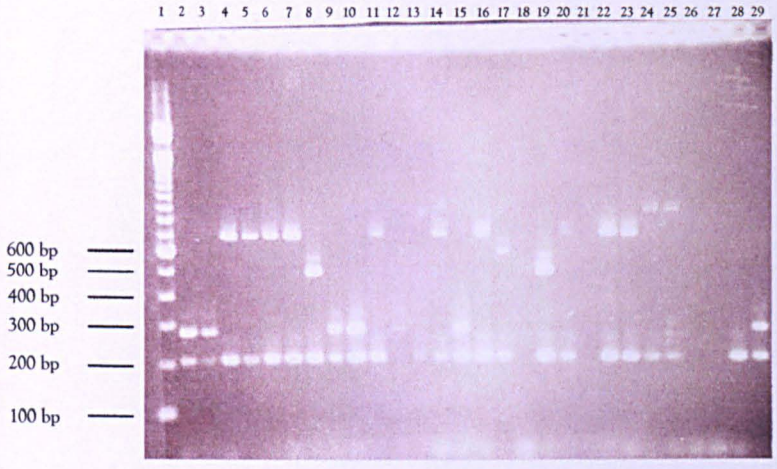
The figure depicts determination of SCC*mec* type I-V based on gel electrophoresis of amplicons obtained from SCC*mec* type I-V multiplex PCR. Lane 1, marker (100 bp ladder, Life technologies); Lanes 2-4, 16, 20, & 24, MSSA; Lanes 5-7, 9, 12-15, 17-18, & 25, SCC*mec* type IV; Lanes 8, 11, & 19, SCC*mec* type II; Lanes 10 & 21, SCC*mec* type V; Lane 22, SCC*mec* type III; Lane 23, SCC*mec* type VI; Lane 26, negative control. The labels on the left indicate the size in base pair (bp) for the 100 bp ladder (Lane 1).

3.2.2.2. SCC*mec* IV subtype

Of the 128 MRSA isolates which belonged to SCC*mec* type IV, 79 isolates (61.7 %) were type IVh, 19 (14.8 %) were IVa and 22 (17.2 %) were non-subtypeable. Subtypes IVb-g

constituted the remaining eight (6.3 %) isolates with one to three isolates assigned to each subtype (Figure 22).

Figure 22. SCC*mec* subtype IVa-h determination based on size variations in SCC*mec* amplicon



The figure depicts determination of SCC*mec* subtype IVa-h based on gel electrophoresis of amplicons obtained from SCC*mec* subtype IVa-h multiplex PCR. Lane 1, marker (100 bp ladder, Life technologies); Lanes 2-3, 9-10, 15, & 29, SCC*mec* type IVa; Lanes 4-7, 11, 14, 16, & 22-23, SCC*mec* type IVh; Lanes 8, & 19, SCC*mec* type IVc; Lane 17, SCC*mec* type IVd; Lanes 24 & 25, SCC*mec* type IVg; Lanes 12-13, & 21, MSSA; Lanes 20 & 28, ccrB2 positive; Lanes 26 & 27, SCC*mec* type VI. The labels on the left indicate the size in base pair (bp) for the 100 bp ladder (Lane 1).

The isolates that could not be subtyped produced a single PCR product for the internal positive control (*ccrB2*). Amongst eight isolates from a single patient, two were non-subtypeable as no PCR product was obtained from the individual PCRs and the remaining six isolates exhibited the presence of SCC*mec* IVa or IVh subtypes.

3.2.2.3. SCC*mec-orfX* junction

Sequence determination of the SCC*mec-orfX* junction was performed to detect the presence of a partial or novel SCC*mec* amongst eight isolates from a single patient. A 1500 bp consensus sequence of the *orfX* and left extremity of SCC*mec* type IV region was generated. PCR primers amplified approximately 1000 bp of this consensus sequence in six of the eight isolates and the two remaining isolates, which were non-typeable for the SCC*mec* type I to V and SCC*mec* type IV multiplex PCRs, produced no amplicon. Three nested primers were designed 100 bp into the amplicon sequence and the two reverse primers were designed 600 and 750 bp into the amplicon sequence. The resultant sequences were assembled into contigs up to 814 bp in size. Greater sequence coverage was obtained in five of the six isolates using the nested primer pairs (Table 16).

Table 16. Sequence coverage of SCC*mec-orfX* junction

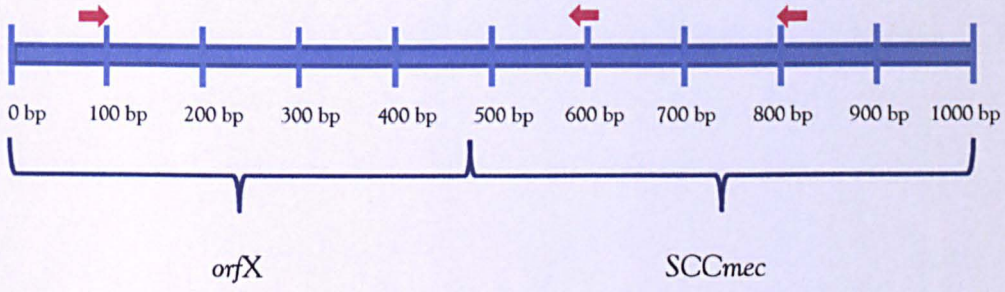
Isolate number	SCC*mec-orfX* PCR 1	SCC*mec-orfX* PCR 2
DBHT 1	399 bp	814 bp
DBHT 2	247 bp	692 bp
DBHT 3	332 bp	807 bp
DBHT 4	N/A*	N/A
DBHT 5	N/A	N/A
DBHT 6	663 bp	805 bp
DBHT 7	725 bp	717 bp
DBHT 8	387 bp	738 bp

Sequence determination of the SCC*mec-orfX* PCR 1 product was performed using PCR primers whilst amplicons from the SCC*mec-orfX* PCR 2 were sequenced using nested primers. Greater sequence coverage was obtained in majority of the isolates using nested primers.

***N/A, not applicable.**

The contigs were aligned against the 1500 bp consensus sequence and BLAST analysis revealed that the contigs covered approximately 480 bp of the *orf*X and 520 bp of the left SCC*mec* type IV region (Figure 23). This confirmed the presence of SCC*mec* type IV in six of the eight isolates and partial or complete SCC*mec* elements were not identified in the two remaining isolates. Alignment of SCC*mec*-*orf*X junction sequences revealed this region was highly conserved among these six isolates. These six isolates belonged to either CC22 or CC672; SNPs, insertions or deletions were not observed in the sequences.

Figure 23. Region depicting the *orfX* and the left extremity of SCC*mec*



Red arrows indicate the annealing sites and direction of amplification for three nested primers utilised to sequence this region.

3.2.2.4. *orf*X amplification

The presence of an intact *orf*X was investigated in eight isolates from a single patient. A consensus sequence sized 480 bp from SCC*mec* type I to V data were used to design primers. A forward primer was designed 24 bp into the consensus sequence whilst a reverse primer was designed at 480 bp. A PCR amplicon approximately 480 bp in length was amplified from the eight isolates.

3.2.2.5. SCC*mec* type VI to VIII

The presence of SCC*mec* types VI to VIII was sought in three isolates from SRL, which were previously non-typeable using the SCC*mec* type I to V and SCC*mec* type IV multiplex PCRs. The three isolates (DBHT 28, 42 and 43) generated amplicons for *mec* class B, *ccrB4* and *ccr* type 4 which corresponded to the presence of SCC*mec* type VI.

3.2.2.6. SCC*mec* non-typeable isolates

The SCC*mec* type present in thirteen isolates from this study were not identified using the above SCC*mec* typing methods. Two multiplex PCRs enabled the assignment of the *ccr* complex type and *mec* class (Kondo *et al.*, 2007). One isolate belonged to SCC*mec* type VI, a combination of *mec* class and *ccr* complex type which did not correspond to a SCC*mec* type were identified in six isolates and a further six isolates which were negative for the presence of the *mecA* gene were assigned as MSSA (Table 17).

3.2.2.7. SCC*mec* data concordance with MLST data

The 184 MRSA isolates belonged to 12 CCs and two singleton STs. The largest number of different SCC*mec* types were identified amongst CC22 and CC8 isolates (Figure 24). One hundred and eighteen MRSA isolates (64 %) were assigned to CC22; 107 (91 %) of the CC22 MRSA isolates belonged to SCC*mec* type IV. Twelve (7 %) of the MRSA isolates belonged to CC30 and 11 (92 %) of these were assigned to SCC*mec* type II. A non-typeable SCC*mec* element was identified in six isolates assigned to CC8, 22 or a singleton ST.

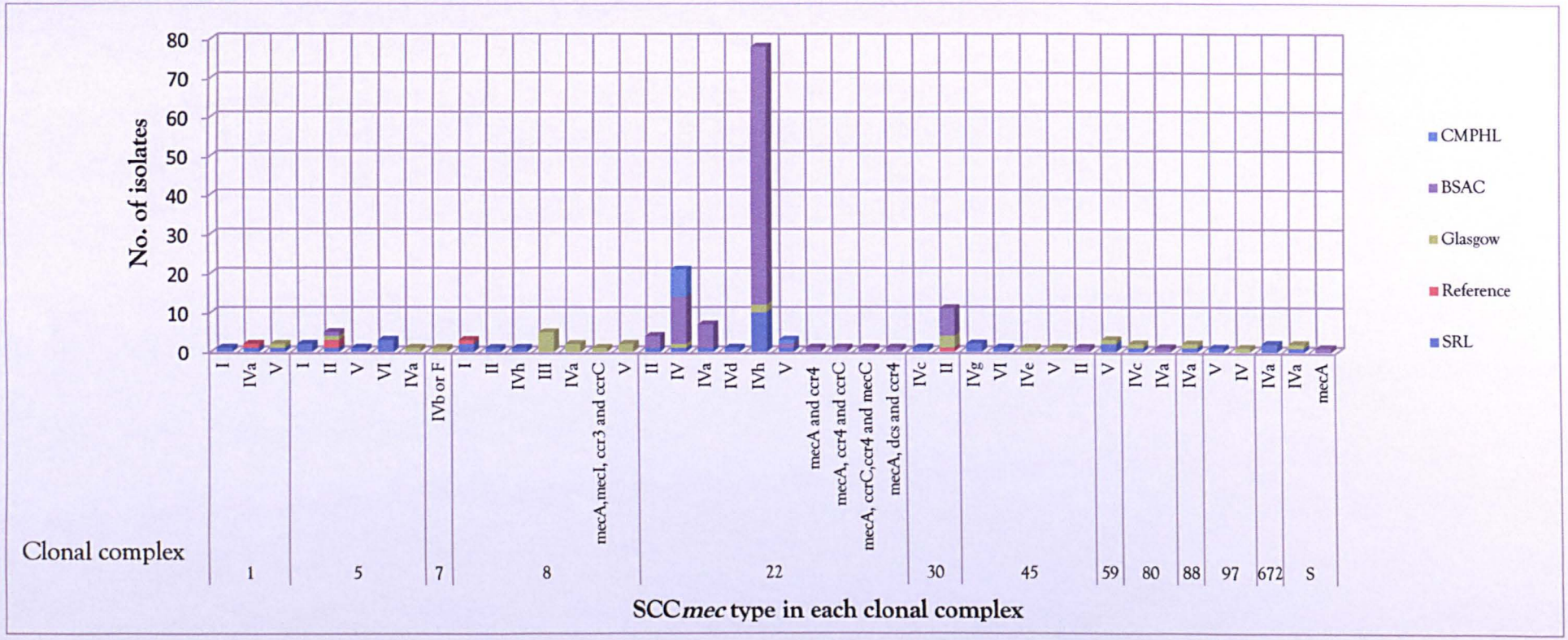
SCC*mec* type IV non-subtypeable elements were detected in isolates belonging to CC22 and 97. The D and CI value was calculated for isolates assigned to the six CCs: CC1, 5, 8, 22, 30 and 45 (Table 18). The calculated CI values overlapped between CC1, 5, 8 and 45. The CI values for CC22 or CC30 overlapped only with those for CC1.

Table 17. Results of SCC*mec* non-typeable isolates

Isolate number	Amplicons	SCC*mec* assignment
DBHT 4	No amplicon	MSSA
DBHT 5	No amplicon	MSSA
DBHT 29	*mec* class B, *mec* internal control, *ccr* type 4	SCC*mec* type VI
DBHT 66	No amplicon	MSSA
DBHT 67	*mecA* internal control, *mecI*, *ccr* type 3 and *ccrC*	SCC*mec* non-typeable
DBHT 81	No amplicon	MSSA
DBHT 117	*ccr* type 4, *mec* internal control	SCC*mec* non-typeable
DBHT 157	*ccrC*, *ccr* type 4, *mec* internal control	SCC*mec* non-typeable
DBHT 163	*mec* internal control	SCC*mec* non-typeable
DBHT 167	*ccr* type 4, *mec* internal control	SCC*mec* non-typeable
DBHT 176	*ccr* type 4, *ccrC*, *mec* internal control and *mec* class C	SCC*mec* non-typeable
DBHT 295	*ccr* type 4	MSSA
DBHT 297	*ccr* type 4	MSSA

PCR amplicons generated from a previously described multiplex PCRs (Kondo *et al.*, 2007).

Figure 24. Distribution of SCC*mec* types in each clonal complex for all isolates in the study



The figure displays the distribution of SCC*mec* types amongst isolates assigned to each clonal complex (CC) or singleton sequence types (S) in relation to one of five isolate collections. CC assignment was based on MLST data.

3.2.2.8. Analysis of SRL collection

Of the 50 SRL isolates, 41 were MRSA and nine MSSA. MRSA isolates belonged to five types of SCC*mec* (I, II and IV-VI) and those which were assigned to SCC*mec* type IV were subtyped into SCC*mec* type IVa, c, d, g, h or were non-subtypeable. The majority of isolates (n=23, 46 %) belonged to SCC*mec* type IV and six isolates (12 %) were assigned to type V. The 41 MRSA isolates belonged to 11 CCs which included CC5, 8, 22, 30 and 45. MRSA isolates assigned to a single CC exhibited one and eight SCC*mec* types (Figure 24). Ten isolates (24 % of MRSA) belonged to ST22-IVh. Majority of the remaining 31 isolates each belonged to a unique CC and SCC*mec* type combination. Twelve of the isolates from this collection were assigned as EMRSA-15 (ST22-MRSA-IV) and none as EMRSA-16 (ST36-MRSA-II). Of the eight isolates from a single patient, two were assigned to ST361-MRSA-IVa, four to ST22-MRSA-IVh and two were MSSA isolates assigned to ST22. Of the 15 isolates assigned to EMRSA-15 by bacteriophage typing and/or PFGE, 12 were confirmed as EMRSA-15 by MLST and SCC*mec* typing, two were ST361 and one was a SLV of ST22.

Table 18. Discriminatory power and confidence intervals based on SCC*mec* data amongst isolates from six clonal complexes

CC	D value	CI value
CC1	0.652	0.404-0.899
CC5	0.788	0.639-0.937
CC8	0.886	0.815-0.956
CC22	0.530	0.433-0.627
CC30	0.167	0.000-0.433
CC45	0.933	0.805-1.000

The discriminatory power (D) (Simpson, 1949) and confidence interval (CI) (Grundmann *et al.*, 2001) values were calculated based on SCC*mec* data amongst isolates obtained from six clonal complexes (CC).

3.2.2.9. Analysis of Reference collection

Five of the six reference strains were confirmed as MRSA and one was an MSSA. The COL strain was assigned to ST250-MRSA-I, the strains N315 and Mu50 were ST5-MRSA-II, MRSA252 belonged to ST36-MRSA-II (EMRSA-16), MW2 was ST1-MRSA-IVa and MSSA476 belonged to ST1-MSSA (Kuroda *et al.*, 2001; Baba *et al.*, 2002; Holden *et al.*, 2004; Gill *et al.*, 2005).

3.2.2.10. Analysis of Glasgow collection

From the Glasgow collection of 34 isolates, 27 were MRSA and seven MSSA. The MRSA and MSSA assignment was concordant with previously performed MICs for oxacillin and *mecA* PCR (performed by SRL). SCC*mec* was not identified amongst four isolates which were

meticillin resistant, based on oxacillin MICs or *mecA* PCR. Twenty-seven MRSA isolates were assigned to four SCC*mec* types (II to V) and one non-typeable element. The majority of MRSA isolates (n=12, 35 %) belonged to SCC*mec* type IV whilst five (15 %) isolates belonged to SCC*mec* type II and V each. The 12 isolates assigned to SCC*mec* type IV were subtyped as IVa-c, IVe, IVh or were non-subtypeable. The three isolates assigned as EMRSA-15 by PFGE were ST22-MRSA-IV. Of the four isolates assigned as EMRSA-16 by PFGE, three were ST36-MRSA-II. The remaining isolate was assigned to ST36 but no SCC*mec* was detected. Amongst MRSA isolates assigned to a single CC, between one and five SCC*mec* types were identified (Figure 24). The remaining isolates (74 %) belonged to a unique CC and SCC*mec* type combination.

3.2.2.11. Analysis of British Society for Antimicrobial Chemotherapy collection

The BSAC collection constituted 99 MSSA isolates and 103 MRSA isolates based on SCC*mec* typing. These results were concordant with previously performed MICs determinations for oxacillin and the *mecA* PCR (BSAC, 2010). The MRSA isolates belonged to one of three SCC*mec* types (12 isolates belonged to type II, 85 were type IV and one was type V) or five were non-typeable. The 85 SCC*mec* type IV isolates were subtyped as IVa (n=7), IVh (n=66) or were non-subtypeable (n=12). The MRSA isolates belonged to five CCs and one singleton ST. Sixty-six MRSA belonged to ST22-IVh and majority of the remaining 37 isolates each belonged to a unique CC and SCC*mec* type combination (Figure 24). The presence of an SCC*mec* type II element was significantly associated withCC30 isolates as compared to CC22 isolates whilst SCC*mec* type IV was significantly associated withCC22 isolates as compared to CC5, 30 and 45.

Of 87 BSAC isolates from 2009, 81 (79 %) isolates were assigned to EMRSA-15 whilst six (6 %) were EMRSA-16. Sixty-four (79 %) EMRSA-15 isolates were assigned to SCC*mec* IVh whilst the remaining isolates belonged to subtype IVa (five isolates) or were non-sub-typeable (12 isolates). Of the 16 (15 %) MRSA isolates not assigned to EMRSA-15 or -16, eight were ST22 (with SCC*mec* type II, V or was non-typeable) and three were SLVs of ST22 (ST954, ST2031 and ST2037 with SCC*mec* type IVa or IVh). Amongst five of the remaining 16 isolates, one was ST5 (with SCC*mec* II), one isolate was ST80 (with SCC*mec* IVa) and one was ST45 (with SCC*mec* II), one was ST779 (with a non-typeable SCC*mec*) and one was a SLV of ST30 (ST1621 with SCC*mec* II).

A higher percentage (≥95 % confidence) of CC22 isolates displayed resistance to ciprofloxacin, oxacillin and piperacillin-tazobactam as compared to isolates assigned to CC5, 8 and 15. Resistance to ciprofloxacin, erythromycin and piperacillin-tazobactam was common amongst the EMRSA-15 and -16 isolates as previously described (Ellington *et al.*,2010; Hope *et al.*, 2008). Statistical analysis based upon antimicrobial resistance and SCC*mec* data were performed on one representative of a specific lineage from each centre to reduce the association between isolates. Antimicrobial resistance and SCC*mec* data were compared between lineages and 95 % CIs were calculated using Stata software using the Clopper-Pearson method (Clopper & Pearson, 1934). However, this reduced the number of isolates belonging to each CC. In addition, one isolate from this subset was chosen at random as a representative of a specific lineage; the results were therefore variable and dependent on the isolate selected. Significant differences could not be identified amongst the other lineages due to broad 95 % CIs owing to the low number of isolates. Resistance to glycopeptides (teicoplanin and vancomycin) was not identified amongst isolates in this study

3.2.2.12. Analysis of Clinical Microbiology and Public Health Laboratory collection

Of the ten isolates from CMPHL, eight were MRSA and two MSSA. Of the nine ST22 isolates, seven belonged to SCC*mec* type IV, one belonged to SCC*mec* type V and one isolate was MSSA. The single ST5 isolate was an MSSA. Of the eight isolates from the ICU outbreak, one belonged to ST5 and the remaining seven isolates were EMRSA-15.

The D value was 0.772 for SCC*mec* data from all isolates with a CI value of 0.720 to 0.823. The D value and CI for isolates from five collections were calculated. The CI values overlapped between the SRL, Reference and Glasgow collection. The CI values also overlapped between the Reference, BSAC and CMPHL collections. The CI values of the SRL and Glasgow collection did not overlap with the CI values of the BSAC and CMPHL collection (Table 19).

Table 19. Discriminatory power and confidence interval based on SCC*mec* data amongst isolates from five isolate collections

Isolate collection	D value	CI value
SRL	0.871	0.821-0.920
Reference	0.700	0.349-1.000
Glasgow	0.892	0.847-0.936
BSAC	0.563	0.458-0.667
CMPHL	0.250	0.000-0.600

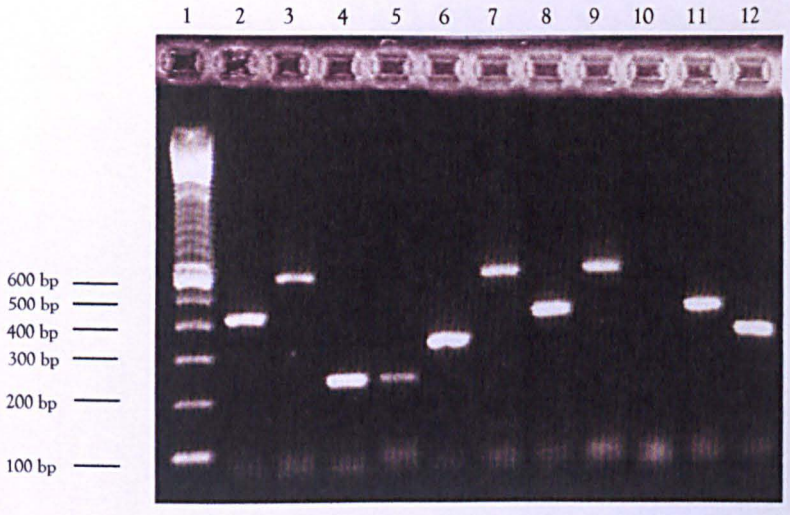
The discriminatory power (D) (Simpson, 1949) and confidence interval (CI) (Grundmann *et al.*, 2001) values were calculated based on SCC*mec* data amongst isolates obtained from each collection.

3.2.3. *spa* typing

Seventy-four MRSA isolates from three collections (SRL, Reference and Glasgow) were investigated using *spa* typing. Only a subset of the isolates were analysed to indicate the level of heterogeneity at this locus and to determine its concordance with MLST data. *spa* amplicons ranging in size from 250 to 900 bp were obtained from 73 isolates which were assigned to 39 *spa* types; the remaining isolate was non-typeable (Figure 25). The most common *spa* types were t032 and t037. Six and five isolates were assigned to *spa* type t032, containing 16 repeats and t037, with seven repeats respectively. The number of repeats in each *spa* type varied from 3 (in t026) to 16 (in t032). Three isolates were assigned to two novel *spa* types; two were t9238 whilst one was t9239 (Appendix VI).

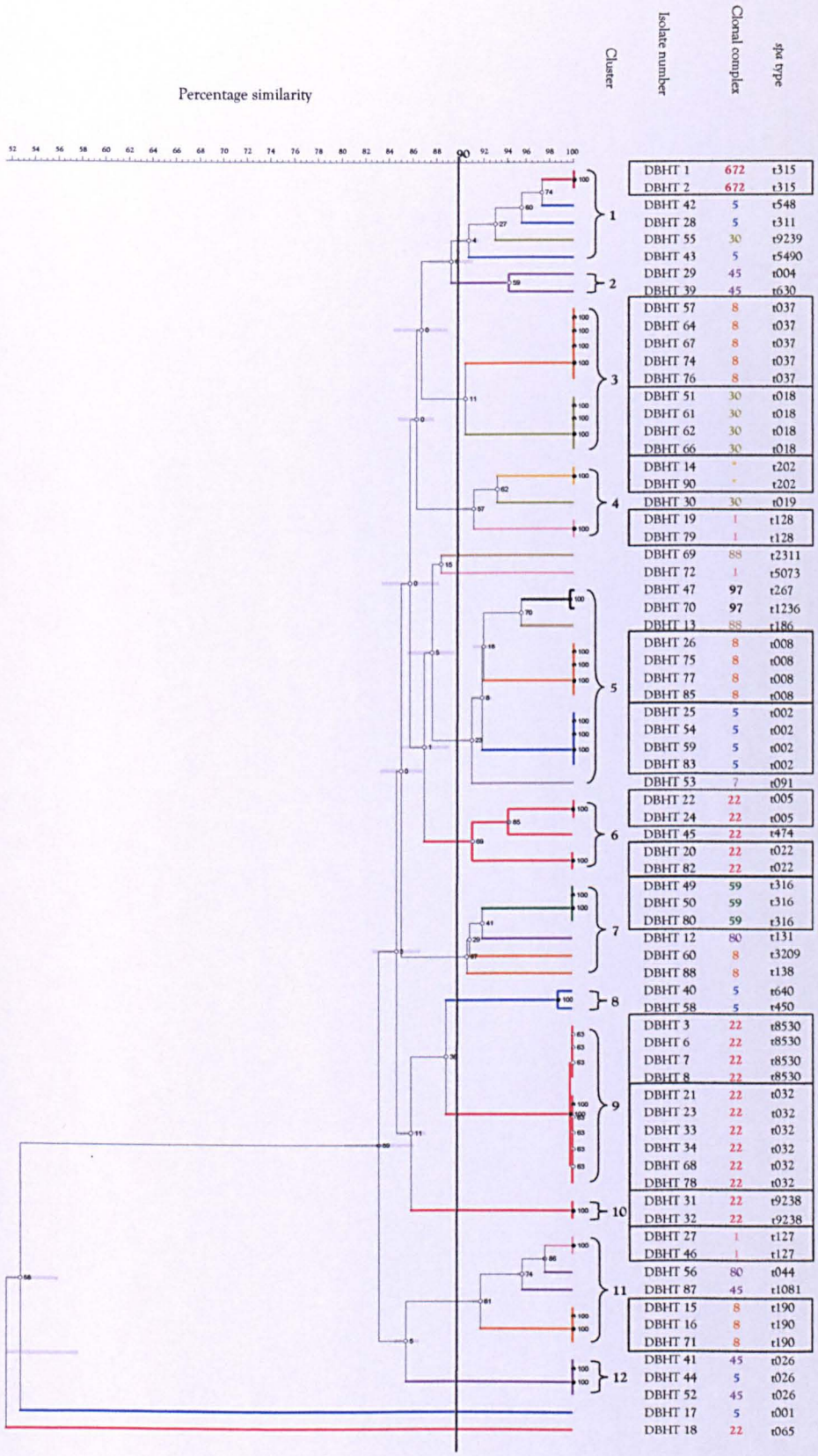
The 73 *spa*-typeable isolates were assigned to 12 CCs and one singleton including CC5, 8, 22, 30 and 45. Forty-three isolates (59 %) belonged to CC5, 8 and 22. Isolates of CC5 and 22 exhibited diverse *spa* types (Figure 26). CC8 and 45 isolates exhibited five and four different *spa* types respectively whilst CC1 and 30 isolates exhibited three different *spa* types each. Isolates assigned to the remaining lineages belonged to up to two different *spa* types. Isolates assigned to the novel *spa* types t9238 and t9239 belonged to CC22 and 30 respectively. Cluster analysis was performed on *spa* sequences using the UPGMA method. Based on MLST data, a 90 % similarity cut-off was applied to the *spa* dendrogram. Twelve *spa* clusters were identified among 69 MRSA isolates and the remaining four isolates exhibited divergent *spa* sequences (>10 % difference) (Figure 26).

Figure 25. *spa* type as determined by size variations in *spa* amplicon



The figure depicts determination of *spa* type based on gel electrophoresis of amplicons obtained from *spa* typing PCR. Lane 1, 100 bp ladder; Lane 2, *spa* type t186; Lane 3,*spa* type t032; Lane 4, *spa* type t026; Lane 5, *spa* type t026; Lane 6, *spa* type t037; Lane 7, *spa* type t032; Lane 8, *spa* type t008; Lane 9, *spa* type t032; Lane 10, *spa* non-typeable; Lane 11, *spa* type t008; Lane 12, *spa* type t1081.

Figure 26. Dendrogram derived from *spa* data using Unweighted Pair Group Method with Arithmetic mean (UPGMA) cluster analysis



Based on 90 % similarity cut-off (shown as a vertical black line), 12 *spa* clusters were identified (shown as black numbers 1 to 12). The clonal complex and *spa* type of isolates is shown on the right of the cluster number. Isolates with identical *spa* profiles and CCs are boxed together.

Cluster 1 grouped six isolates which belonged to CC672, 5 and 30 and exhibited five different *spa* types. The two isolates in cluster 2 and cluster 8 were assigned to CC45 and 5 respectively; each isolate belonged to unique *spa* types. Of the nine isolates in cluster 3, five belonged to CC8 and exhibited *spa* type t037 and four isolates belonged to CC30 and exhibited *spa* type t018. Cluster 4 consisted of five isolates; two belonged to CC1 and exhibited *spa* type t128, one to CC30 exhibited *spa* type t019, and two to the singleton ST93 exhibited *spa* type t202. The isolates within cluster 5 belonged to five CCs, two CC97 isolates belonged to *spa* type t267 and t1236 whilst isolates assigned to the four CCs exhibited four *spa* types. Clusters 6, 9 and 10 consisted of isolates belonging to CC22 which exhibited six distinct *spa* types. Of the six isolates within cluster 7, three isolates were CC59 and exhibited *spa* type t316 whilst the remaining three isolates (CC8 or CC80) exhibited individual *spa* types. The seven isolates in cluster 11 belonged to diverse CCs (CC1, 8, 45 and 80). The two CC1 isolates had *spa* type t127 and the three CC8 isolates belonged to t190; the individual isolates belonging to CC45 and 80 exhibited individual *spa* types. The three isolates in cluster 12 were assigned to CC5 or 45 and belonged to the *spa* type t026. Four isolates belonging to CC1, 5, 22 and 88 exhibited divergent *spa* sequences.

The D and CI values based on *spa* data were calculated for isolates from six CC; CI values overlapped between the six CCs (Table 20).

Table 20. Discriminatory power and confidence intervals of *spa* data from isolates of six clonal complexes

CC	D value	CI value
CC1	0.800	0.657-0.943
CC5	0.891	0.733-1.000
CC8	0.791	0.689-0.892
CC22	0.843	0.744-0.941
CC30	0.600	0.215-0.985
CC45	0.900	0.724-1.000

The discriminatory power (D) (Simpson, 1949) and confidence interval (CI) (Grundmann *et al.*, 2001) values were calculated based on *spa* data amongst isolates obtained from six clonal complexes (CC).

3.3. Genetic diversity in the whole genome

3.3.1. Preliminary fluorescent amplified fragment length polymorphism (FAFLP) analysis

FAFLP was utilised to assess the degree of genetic diversity between isolates belonging to the same CC. DBHT 1 to 30 were analysed using four endonuclease combinations: *Eco*RI with *Mse*I, *Hin*dIII with *Hha*I, *Eco*RI with *Hha*I and *Bgl*II with *Csp*6I. Various selective and non-selective primer combinations were utilised for amplification of the restricted fragments. FAFLP analysis using the primer combinations *Hin*dIII+A and *Hha*I+T, *Eco*RI+0 and *Hha*I+A, *Bgl*II+0 and *Csp*6I+T, and *Bgl*II+0 and *Csp*6I+C was successful for all isolates analysed. FAFLP using the remaining three primer combinations (*Eco*RI+0 and *Mse*I+C, *Bgl*II+0 and *Csp*6I+A, and *Bgl*II+0 and *Csp*6I+G) was successful for up to 26 of the 30 isolates. The remaining isolates generated profiles with less than 34 AFs which indicated a

failure in FAFLP digestion, ligation or in the PCR reaction. Therefore FAFLP using these three primer combinations was not successful for all isolates.

The number of AFs in each profile indicated the number of loci sampled in the genome. Chi-squared analysis of the total number of AF markers, the number of markers common to all profiles and the number of polymorphic markers amongst profiles differed significantly amongst each of the primer combinations (99.99 % confidence). Among the profiles generated with the different primer combinations, profiles produced with *Eco*RI+0 and *Mse*I+C were composed of the least number of polymorphic or total AF markers. This primer combination sampled fewer loci and therefore was likely to display less discriminatory power. The largest number of AFs and those that were common or polymorphic were generated using the primers *Eco*RI+0 and *Hha*I+A. Although the discriminatory power of a technique increases as the number of loci sampled increases the possibility of errors due to factors such as homoplasy and collision also increase (Gort and van Eeuwijk, 2012). Therefore a primer combination that generated profiles displaying sufficient discriminatory power was sought. All profiles generated with each of the primer combinations displayed a greater number of fragments between 50 to 300 bp than between 300 to 600 bp and each FAFLP profile covered between 0.40 to 1.33 % of the genome (Table 21). Profiles were required which displayed an approximately uniform spread of a large number of AFs to reduce the chances of errors due to homoplasy and collision.

Comparative analysis of AFs generated from each of the different primer combinations was performed amongst isolates belonging to one of seven major CCs (CC1, 5, 8, 22, 30, 45 and 59). The presence of each AF was classified as follows: a) AFs present in all isolates, b) AFs present in all isolates of one or more lineages, c) AFs present exclusively in isolates of a

specific lineage or d) AFs polymorphic amongst isolates of one or more lineages (Table 22). Factors relating to the success of a particular lineage may be encoded on regions which generated lineage-specific AFs. Lineage-specific AFs were defined as AFs present in ≥95 % of isolates of a single lineage and present in <5 % of the isolates of other lineages. Using these criteria, lineage-specific AFs were identified using all primer combinations with the exception of *Eco*RI+0 with *Mse*I+C. The primer combinations *Bgl*II+0 with *Csp*6I+T and *Bgl*II+0 with *Csp*6I+C identified the greatest number of CC1-specific AFs. Profiles generated with the primer set *Bgl*II+0 with *Csp*6I+C displayed the largest number of CC5 and 22-specific AFs. *Bgl*II+0 with *Csp*6I+A profiles generated the largest number of lineage-specific AFs amongst CC8 isolates. From this initial analysis it was concluded the *Bgl*II+X and *Csp*6I+X primer combination displayed reasonable discriminatory power between lineages whilst reducing the possibility of errors to homoplasy and collision. However as multiple *Bgl*II+0 and *Csp*6I+X primer combinations displayed reduced typeability a larger set of isolates were examined using numerous *Bgl*II+X and *Csp*6I+0 primer combinations.

Isolates 1 to 50 from SRL were examined with the *Bgl*II+X and *Csp*6I+0 primer combination. FAFLP with the primer combinations *Bgl*II+T with *Csp*6I+0 and *Bgl*II+A with *Csp*6I+0 was successful for all isolates analysed. Of the 50 isolates analysed using the primers *Bgl*II+C with *Csp*6I+0 and *Bgl*II+G with *Csp*6I+0, 14 and 20 isolates respectively produced profiles with peaks of low signal intensity and were eliminated from further analysis. Profiles generated using the primers *Bgl*II+A and *Csp*6I+0 displayed an approximately uniform spread of AFs between 50 to 600 bp. Conversely, majority of the AFs (84 %) within profiles generated using the primers *Bgl*II+C and *Csp*6I+0 were <300 bp (Table 21).

Table 21. FAFLP profile characteristics with different primer combinations

Primer combination*	No. of AFs in each profile†	No. of AFs sized 50-300 bp (%)†	No. of AFs sized 300-600 bp (%)†	No. of common AFs‡	No. of polymorphic AFs‡	No. of AFs‡	Profile coverage (%)§
*Bgl*II+0 and *Csp*6I+A	66 to 94	71	29	20	228	248	0.69 %
*Bgl*II+0 and *Csp*6I+C	53 to 105	65	35	19	277	296	0.62 %
*Bgl*II+0 and *Csp*6I+G	55 to 99	73	27	16	254	270	0.67 %
*Bgl*II+0 and *Csp*6I+T	64 to 119	61	39	20	294	314	0.94 %
*Bgl*II+A and *Csp*6I+0	47 to 123	56	44	14	256	270	0.56 %
*Bgl*II+C and *Csp*6I+0	50 to 147	84	16	7	307	314	0.56 %
*Bgl*II+G and *Csp*6I+0	40 to 85	67	33	9	255	264	0.46 %
*Bgl*II+T and *Csp*6I+0	59 to 131	74	26	24	285	309	0.61 %
*Eco*RI+0 and *Hha*I+A	106 to 159	57	43	35	335	370	1.33 %
*Eco*RI+0 and *Mse*I+C	30 to 56	63	37	11	126	137	0.40 %
*Hin*dIII+A and *Hha*I+T	68 to 137	58	42	19	335	354	0.96 %

*Multiple primer combinations were used in preliminary FAFLP experiments to assess the suitability of the endonuclease and primer pair on all study isolates.

†Indicates number of amplified fragments (AFs) and the distribution of AFs in FAFLP profiles generated from one of eleven primer combinations.

‡The total number of AFs and the number of these which were common (present ≥95 % of isolates) or polymorphic (present <5 % of isolates) amongst all isolates.

§The percentage of the genome sampled using each primer combination.

Table 22. Number of AFs identified for each primer combination amongst seven clonal complexes

FAFLP PCR		Clonal complex						
		CC1	CC5	CC8	CC22	CC30	CC45	CC59
*Bgl*II+0 and *Csp*6I+A	Common AFs	-	-	60(20)	48(20)	-	-	-
	Lineage-specific AFs	-	-	3	4	-	-	-
	Polymorphic AFs	-	-	47	99	-	-	-
*Bgl*II+0 and *Csp*6I+C	Common AFs	45(19)	51(19)	44(19)	43(19)	-	-	-
	Lineage-specific AFs	4	3	1	9	-	-	-
	Polymorphic AFs	38	65	80	141	-	-	-
*Bgl*II+0 and *Csp*6I+G	Common AFs	-	52(16)	60(16)	48(16)	-	-	-
	Lineage-specific AFs	-	2	2	4	-	-	-
	Polymorphic AFs	-	59	56	85	-	-	-
*Bgl*II+0 and *Csp*6I+T	Common AFs	80(20)	76(20)	67(20)	61(20)	-	-	-
	Lineage-specific AFs	4	2	1	4	-	-	-
	Polymorphic AFs	38	64	59	90	-	-	-
*Bgl*II+A and *Csp*6I+0	Common AFs	22(14)	19(14)	20(14)	19(14)	16(14)	19(14)	29(14)
	Lineage-specific AFs	0	1	1	1	3	6	6
	Polymorphic AFs	9	26	61	18	24	14	7
*Bgl*II+C and *Csp*6I+0	Common AF's	39(7)	32(7)	28(7)	30(7)	35(7)	36(7)	-
	Lineage-specific AFs	1	0	1	5	1	1	-
	Polymorphic AFs	61	140	133	185	112	99	-
*Bgl*II+G and *Csp*6I+0	Common AFs	34(9)	26(9)	32(9)	30(9)	28(9)	-	-
	Lineage-specific AFs	2	1	0	6	5	-	-
	Polymorphic AFs	41	87	88	88	48	-	-

Table 22 continued. Number of AFs identified for each primer combination amongst seven clonal complexes

FAFLP PCR		Clonal complex						
		CC1	CC5	CC8	CC22	CC30	CC45	CC59
*Bgl*II+T and *Csp*6I+0	Common AFs	58(24)	48(24)	39(24)	45(24)	56(24)	-	62(24)
	Lineage-specific AFs	1	1	0	4	2	-	2
	Polymorphic AFs	49	67	119	97	59	-	19
*Eco*RI+0 and *Hha*I+A	Common AFs	97(35)	90(35)	107(35)	86(35)	-	-	-
	Lineage-specific AFs	2	2	1	5	-	-	-
	Polymorphic AFs	48	82	57	158	-	-	-
*Eco*RI+0 and *Mse*I+C	Common AFs	16(11)	18(11)	-	26(11)	-	-	-
	Lineage-specific AFs	0	0	-	0	-	-	-
	Polymorphic AFs	62	47	-	88	-	-	-
*Hin*dIII+A and *Hha*I+T	Common AFs	62(19)	56(19)	62(19)	55(19)	-	-	-
	Lineage-specific AFs	0	1	0	4	-	-	-
	Polymorphic AFs	72	89	81	185	-	-	-

Number of amplified fragments (AFs) common to isolates of each lineage, polymorphic amongst isolates of each lineage and specific to isolates of each lineage identified amongst profiles from each of eleven FAFLP primer combinations are displayed. Numbers in brackets indicate the number of AFs common to ≥95 % of all isolates.

The greatest number of lineage-specific AFs were identified amongst profiles generated using the primers *Bgl*II+A and *Csp*6I+0 (n=18), although *Bgl*II+G with *Csp*6I+0 profiles generated the largest number of lineage-specific AFs (n=5) amongst CC30 isolates. Amongst those isolates assigned to CC45 and 59, the greatest number of lineage-specific AFs (n=6) were

identified from *Bgl*II+A with *Csp*6I+0 profiles. Of all lineage-specific AFs identified with each primer combination, the largest proportion of AFs (up to 80 %) were specific to CC22 in the majority of combinations (Table 22). These findings indicated the *Bgl*II+A and *Csp*6I+0 primer combination identified the greatest degree of heterogeneity between lineages.

3.3.1.1. Cluster analysis

Eleven dendrograms were generated for data from each of the primer combinations (data not shown). Based on FAFLP data from each primer combination with the exception of *Bgl*II+A and *Csp*6I+0 data, each isolate exhibited a unique profile. Based on MLST data, a 70 % similarity cut-off was applied to the dendrograms (Table 23).

Table 23. Summary of FAFLP dendrograms based on 11 primer combinations

FAFLP PCR	No. of clusters	No. of divergent profiles	Clonal complex							
			CC1	CC5	CC8	CC22	CC30	CC45	CC59	CC672
*Eco*RI+0 and *Mse*I+C	8	4	N/A*	N/A	N/A	N/A	N/A	N/A	N/A	N/A
*Eco*RI+0 and *Hha*I+A	5	12	26	37	28	N/A	36	42.5	N/A	17
*Hind*III+A and *Hha*I+T	5	20	N/A	50	49	59	52	N/A	N/A	46
*Bgl*II+0 and *Csp*6I+A	4	10	19	42	36	51	40	N/A	N/A	21
*Bgl*II+0 and *Csp*6I+T	5	13	24	36	40	50	38	N/A	N/A	21

Table 23 continued. Summary of FAFLP dendrograms based on 11 primer combinations

FAFLP PCR	No. of clusters	No. of divergent profiles	Clonal complex							
			CC1	CC5	CC8	CC22	CC30	CC45	CC59	CC672
*Bgl*II+0 and *Csp*6I+C	5	17	41	45	28	48	47	N/A	N/A	24
*Bgl*II+0 and *Csp*6I+G	3	16	27	42	37	35	33	N/A	N/A	42
*Bgl*II+T and *Csp*6I+0	8	16	33	39	30	45	33	31	18	43
*Bgl*II+C and *Csp*6I+0	3	30	47.5	59	N/A	58	58	61	55	N/A
*Bgl*II+G and *Csp*6I+0	1	26	44	44	N/A	52	51	52	N/A	31
*Bgl*II+A and *Csp*6I+0	11	8	20	36	33	29	34	35	18	0

Based on a 70 % similarity cut-off, the numbers of clusters and divergent profiles amongst dendrograms generated from one of 11 primer combinations is displayed. In addition, the percentage divergences displayed amongst isolates of each of eight clonal complexes are shown.

'N/A indicates only single representative of a clonal complex present or isolates of the same clonal complex not found in a single cluster.

Dendrograms based on data from primer combinations *Bgl*II+A with *Csp*6I+0 and *Bgl*II+T with *Csp*6I+0 grouped isolates belonging to each of the major lineages. Therefore the divergence between each of these lineages was identified based on these two dendrograms. In addition, the dendrogram based on *Bgl*II+A and *Csp*6I+0 data exhibited the greatest number

of clusters and the least number of divergent profiles (>30 %). Based on the findings of cluster analysis and the initial results of FAFLP profiles, FAFLP analysis was performed on all isolates using the primer combination *Bgl*II+A and *Csp*6I+0.

3.3.2. FAFLP analysis

3.3.2.1. *Bgl*II+A and *Csp*6I+0 FAFLP analysis

FAFLP analysis using the primers *Bgl*II+A with *Csp*6I+0 was performed on all isolates in this study. FAFLP profiles of the 302 isolates exhibited between 35 and 123 AFs. Based on reproducibility studies, profiles from the same isolate displayed >98 % similarity. Of the AFs in each profile, 59 % of the AFs were less than 300 bp in size and 41 % were between 300 and 600 bp based on the mean. The mean genome size calculated from 19 WGSs was 2849090 bp; based on this, each profile covered approximately 0.54 % of the genome.

Amongst the 302 isolates in this study, a total of 512 precisely-sized AFs were identified. The majority of AFs (500 AFs, 98 %) were polymorphic among the profiles from 302 isolates. The remaining 12 AFs (C1-C12) were common to ≥95 % isolates (Table 24). Fragments similar in size to the common AFs were identified via *in silico* analysis of WGSs. Three of the common AFs (C1, C6 and C12) could not be mapped to WGSs due to the presence of multiple AFs within ±3 bp of the expected AF size amongst *in silico* profiles. *In silico* analysis of one other AF (F9) revealed the absence of AFs sized within ±3 bp of the expected size. Therefore the sequence and function of these four AFs could not be discerned. The sequences and putative functions of the remaining eight common AFs were determined via *in silico* analysis. BLAST search results of these eight sequences revealed they encoded essential cell functions. The sequence for fragment C2 partially encoded the *codY* gene which

acts as a transcriptional repressor. Fragment C3 encoded an enzyme involved in the catabolism of glycerol whilst C10 encoded an enzyme required for the metabolism of pyruvate. The sequence of C7 encoded a protein essential for the transfer of a phosphate group to and from adenosine triphosphate. C4 and C8 encoded enzymes required for cell division or nitrogen fixation in other species of bacteria.

The percentage similarity of each of the eight AF sequences (C2-C5, C7-C8, C10-C11) was calculated amongst 10 WGSs. Seven of the eight sequences displayed 97 % to 100 % conservation. One sequence (C5) displayed 77 % similarity amongst the 10 WGSs and represented the region between two ORFs. With the exception of this sequence, the remaining seven sequences were within an ORF.

A search was performed for AFs present in ≥95 % of all MRSA or MSSA isolates and <5 % of the remaining MSSA or MRSA isolates respectively. However, no AFs were exclusive to all FAFLP profiles of MRSA or MSSA isolates. AFs present in FAFLP profiles of the majority of isolates (≥95 %) from each collection were identified as common AFs. Those AFs which were present in the majority of isolates from one collection and in <5 % of the remaining isolates were identified as AFs exclusive to a specific isolate collection. The remaining AFs amongst isolates from each collection were termed polymorphic as differences in the presence or absence of these AFs are due to sequence variations. Amongst isolates obtained from the five collections 19, 12, 13, 21 and 16 common AFs (including the 12 AFs common to all *S. aureus*) were identified within isolates obtained from SRL, BSAC, Glasgow, Reference and CMPHL collection respectively. No AFs were exclusive to isolates from a single collection.

Table 24. AFs common amongst the 302 isolates in this study

AF* number	Size of AF (bp)*	Presence in isolates (%)†	*In silico* AF size‡	BLAST search result of sequence of AF§
C1	65	99.01	N/A‖	N/A
C2	79	99.01	58	Transcriptional repressor CodY
C3	114	98.01	92	Dihydroxyacetone kinase subunit DhaK
C4	152	99.34	129	DNA translocase (FtsK/SpoIIIE family)
C5	160	97.35	140	Region between two open reading frames
C6	189	99.01	N/A	N/A
C7	227	98.01	204	Phosphoglycerate kinase pgk
C8	254	99.01	231	Aminotransferase NifS
C9	281	98.34	N/A	N/A
C10	313	96.36	292	2-oxoisovalerate dehydrogenase
C11	314	98.68	294	Hypothetical protein
C12	392	95.03	N/A	N/A

*Size of amplified fragments (AFs), generated using FAFLP analysis with primers *Bgl*I+A and *Csp*6I+0, common to ≥95 % of isolates.

†Prevalence of common AF amongst all isolates.

‡Size of AF detected via *in silico* analysis in ALFIE (http://www.hpa-bioinformatics.org.uk/cgi-bin/ALFIE/index.cgi).

§BLAST search result of common AF sequences in the NCBI database (http://www.ncbi.nlm.nih.gov/).

‖N/A, not applicable.

3.3.2.2. Reproducibility studies

The reproducibility of the assay was evaluated by performing FAFLP analysis in duplicate on 18 isolates. FAFLP analysis was performed from the same DNA preparation and from two different DNA preparations. FAFLP profiles generated from the same DNA preparation displayed 100 % identity. FAFLP profiles from different DNA preparations of the same

isolate exhibited ≥98 % similarity. Based on these results, FAFLP profiles were considered to be indistinguishable when <3 AF differences (≥98 % similarity) were exhibited. Based on this criterion, 281 profiles were identified amongst the 302 isolates.

3.3.2.3. Concordance of experimental FAFLP profiles with *in silico* profiles

The percentage concordance between *in silico* profiles and the six reference strains was calculated. Between 55 and 70 AFs were exhibited within profiles generated from *in silico* analysis whilst 46 to 59 AFs were identified within experimental profiles. The isolates displayed between 84 and 89 % concordance with *in silico* profiles (Table 25). In each of the six comparisons, an absence of a subset of AFs expected from *in silico* analysis in the experimental profiles accounted for the dissimilarity between profiles.

3.3.2.4. Cluster analysis

A dendrogram was generated of all 302 *S. aureus* isolates based on *Bgl*II+A and *Csp*6I+0 FAFLP data (Appendix VII). Two hundred and eighty-one profiles were identified amongst the isolates. Based on MLST data, a 65 % similarity cut-off was applied to the dendrogram. Twenty-one clusters were exhibited amongst 288 isolates and the remaining 14 isolates displayed divergent profiles (>35 % difference) (Table 26). Cluster 1 consisted of two isolates displaying unique profiles or 17.5 % divergence. The two isolates exhibited 59 AFs, 42 AFs were common whilst 17 were polymorphic. The two MSSA isolates represented two singleton STs.

Table 25. Concordance of experimental FAFLP profiles with *in silico* profiles

Strain name	Isolate number	No. of AFs* absent in experimental profiles	Similarity (%)
MRSA252	DBHT 51	7	89
N315	DBHT 54	8	87
Mu50	DBHT 59	11	84
COL	DBHT 75	8	86
MW2	DBHT 79	9	84
MSSA476	DBHT 84	9	84

Concordance of experimental FAFLP profiles generated using primers *Bgl*II+A and *Csp*6I+0 with *in silico* profiles generated in ALFIE (http://www.hpa-bioinformatics.org.uk/cgi-bin/ALFIE/index.cgi).

*AF, amplified fragments.

The nine isolates grouped in cluster 2 exhibited distinct profiles with between 14.5 and 34.5 % divergence. A total of 203 AFs were identified; 45 were common in all isolates whilst 158 were variable amongst isolates. Of the nine isolates in cluster 2, three were MSSA and a non-subtypeable SCC*mec* type IV was identified in six isolates. Cluster 2 represented 7 % of CC22 isolates in this study (Figure 27).

Table 26. Characteristics of Unweighted Pair Group Method with Arithmetic mean-derived dendrogram based on *Bgl*II+A and *Csp*6I+0 FAFLP data

Cluster no. (no. of isolates)	Divergence (%)*	No. common AFs†	No. polymorphic AFs†	CCs in cluster‡	No. MSSA isolates
1 (n=2)	17.5	42	17	0 (2 singleton STs)	2
2 (n=9)	34.5	45	158	CC22	3
3 (n=117)	33.0	42	136	CC22	6
4 (n=18)	35.0	39	110	CC5	9
5 (n=3)	22.0	44	25	CC5	0
6 (n=10)	28.5	30	53	CC1	6
7 (n=6)	33.5	28	62	CC1 and 25 (and one singleton ST)	4
8 (n=2)	32.0	36	33	CC6 and 101	2
9 (n=4)	26.5	30	30	CC7	3
10 (n=16)	28.0	38	87	CC8	13
11 (n=2)	0.0	50	0	CC672	0
12 (n=3)	33.5	39	40	CC80	0
13 (n=12)	33.0	24	97	CC8	1
14 (n=2)	35.0	46	37	0 (one singleton ST)	0
15 (n=8)	25.5	39	47	CC59	5
16 (n=35)	30.5	31	135	CC30	23
17 (n=3)	31.5	38	37	CC97	1
18 (n=3)	30.5	37	40	CC45	1
19 (n=17)	28.0	37	88	CC45	13
20 (n=2)	38.0	59	57	CC45	2
21 (n=14)	33.0	32	62	CC9, 15, 20 and 121	14

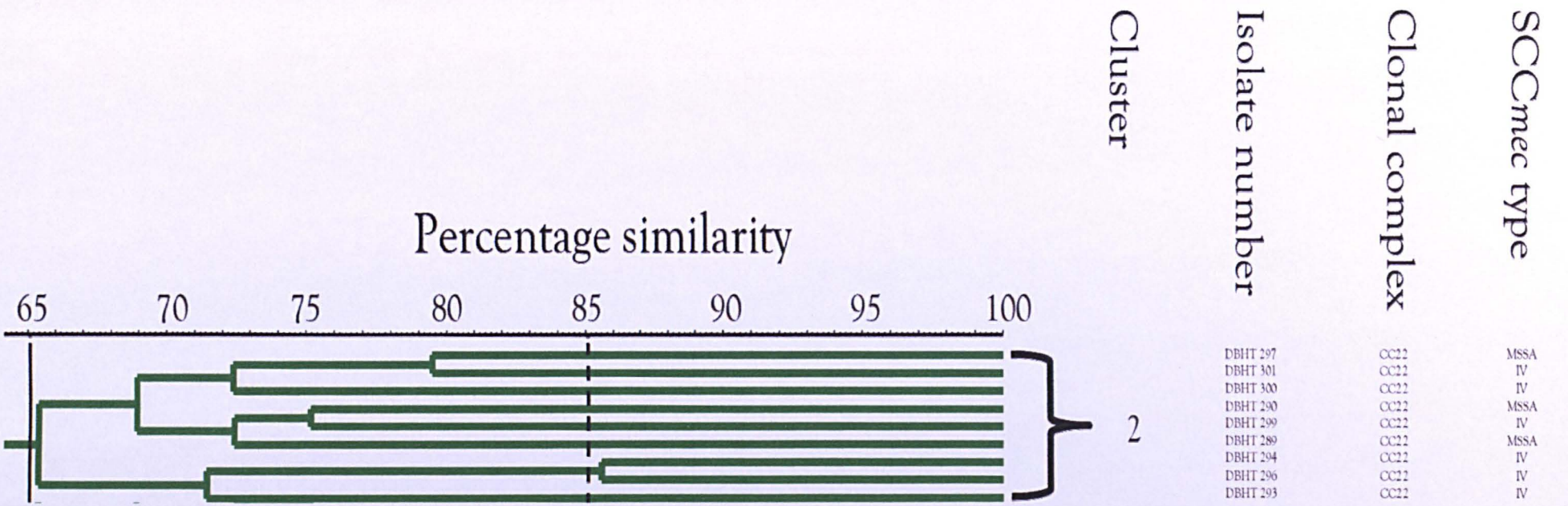
The table displays characteristics for clusters (based on 65 % similarity cut-off) generated from a dendrogram based on *Bgl*II+A and *Csp*6I+0 FAFLP data.

*The percentage divergence exhibited between profiles within each cluster is displayed.

†The number of amplified fragments (AFs) common to ≥95 % of isolates and <95 % of isolates (polymorphic) within each cluster are displayed.

‡The clonal complex (CC) assignment of isolates within each cluster is displayed and those assigned to singleton sequence types (STs) are shown in brackets.

Figure 27. Cluster 2 of UPGMA-derived dendrogram based on *Bgl*II+A and *Csp*6I+0 FAFLP data



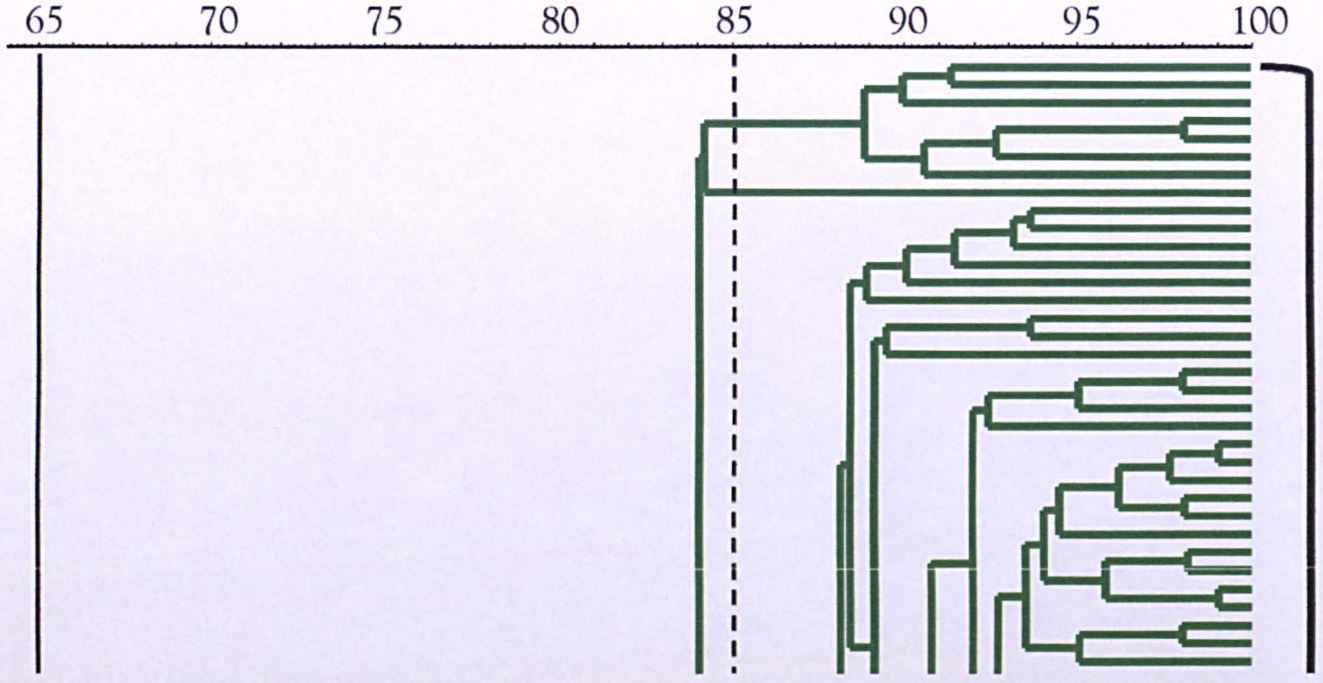
Cluster 2 was based on 65 % similarity cut-off (shown as vertical black line). The dashed line represents 85 % similarity cut-off. The clonal complex (CC) and SCC*mec* type of each isolate is shown on the right.

One hundred and seventeen isolates constituting cluster 3 showed one of 98 profiles which exhibited between 0 to 33.0 % divergence. Forty-two AFs were conserved amongst the profiles and 136 AFs were different. Ninety-one percent (n=117) of the CC22 isolates in this study were grouped in Cluster 3. Of these isolates, 111 belonged to one of five SCC*mec* types and six isolates were MSSA (Figure 28).

Cluster 4 was composed of 18 isolates which were assigned to 17 profiles which exhibited between 1 to 35 % divergence. Thirty-nine of the 149 AFs were common to profiles of all isolates in this cluster and 110 AFs were polymorphic. Seventy-eight percent (n=18) of CC5 isolates constituted cluster 4; nine isolates were assigned to one of four SCC*mec* types and the remaining nine were MSSA (Figure 29). Three isolates amongst cluster 5 had unique profiles which exhibited up to 22 % divergence. The isolates displayed 44 common AFs and 25 variable AFs. Cluster 5 consisted of 13 % of CC5 isolates which belonged to SCC*mec* type II or IVa (Figure 29).

Cluster 6 comprised 10 isolates displaying distinct profiles with ≤28.5 % divergence. Thirty AFs were conserved amongst the isolates whilst 53 were polymorphic. Eighty-three percent (n=10) of CC1 isolates constituted cluster 6, four isolates exhibited one of three SCC*mec* types and six were MSSA (Figure 30). Each of the six isolates in cluster 7 displayed a unique profile with up to 33.5 % divergence. Profiles shared 28 AFs whilst 62 AFs were variable. Two isolates amongst cluster 7 were assigned to CC1, three to CC25 and one was a singleton ST. One CC1 and the CC25 isolates were MSSA and an SCC*mec* type V or a non-typeable element were identified in the remaining two isolates (Figure 30).

Figure 28. Cluster 3 of UPGMA-derived dendrogram based on *Bgl*II+A and *Csp*6I+0 FAFLP data



Cluster	Isolate number	Clonal complex	SCC*mec* type
	DBHT 187	CC22	IVa
	DBHT 191	CC22	IVa
	DBHT 199	CC22	MSSA
	DBHT 188	CC22	IVa
	DBHT 189	CC22	II
	DBHT 192	CC22	IVa
	DBHT 193	CC22	IVa
	DBHT 182	CC22	IVa
	DBHT 123	CC22	IVa
	DBHT 167	CC22	-
	DBHT 184	CC22	IVa
	DBHT 131	CC22	IVa
	DBHT 213	CC22	MSSA
	DBHT 179	CC22	IVa
	DBHT 112	CC22	IVa
	DBHT 116	CC22	IVa
	DBHT 110	CC22	IV
	DBHT 141	CC22	IVa
	DBHT 146	CC22	IVa
	DBHT 145	CC22	IV
	DBHT 157	CC22	-
	DBHT 168	CC22	IVa
	DBHT 183	CC22	IVa
	DBHT 172	CC22	IVa
	DBHT 136	CC22	IVa
	DBHT 159	CC22	IVa
	DBHT 128	CC22	IV
	DBHT 125	CC22	IVa
	DBHT 153	CC22	IVa
	DBHT 152	CC22	IVa
	DBHT 154	CC22	IVa
	DBHT 150	CC22	IVa
	DBHT 151	CC22	IVa
	DBHT 115	CC22	IVa
	DBHT 130	CC22	IVa

Figure 28 continued. Cluster 3 of UPGMA-derived dendrogram based on *Bgl*II+A and *Csp*6I+0 FAFLP data

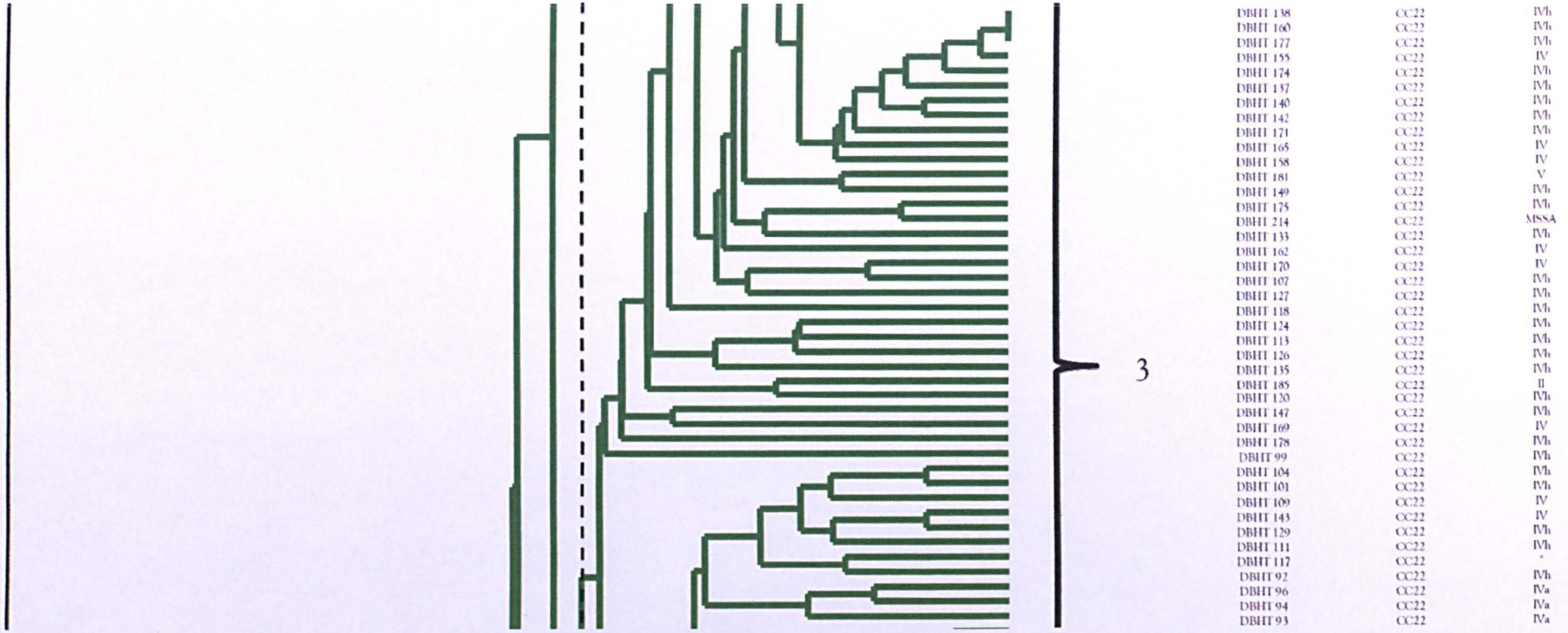
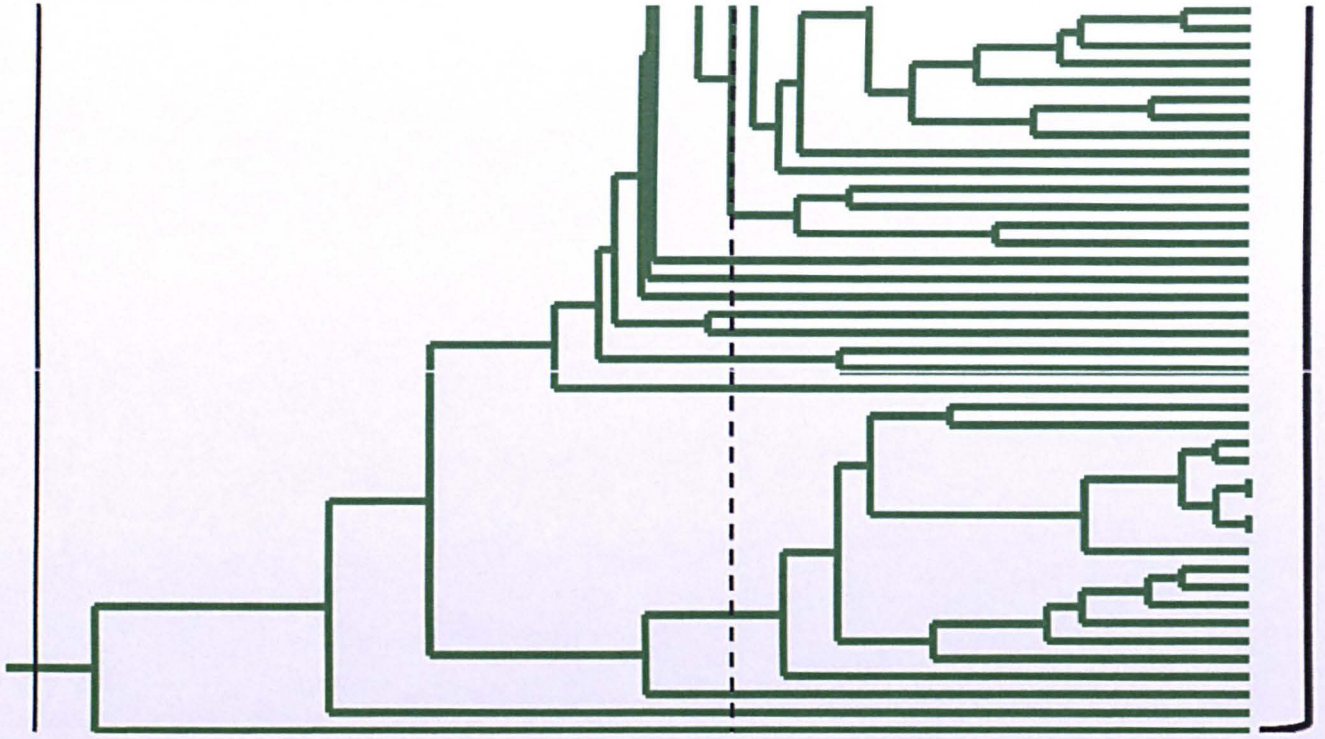


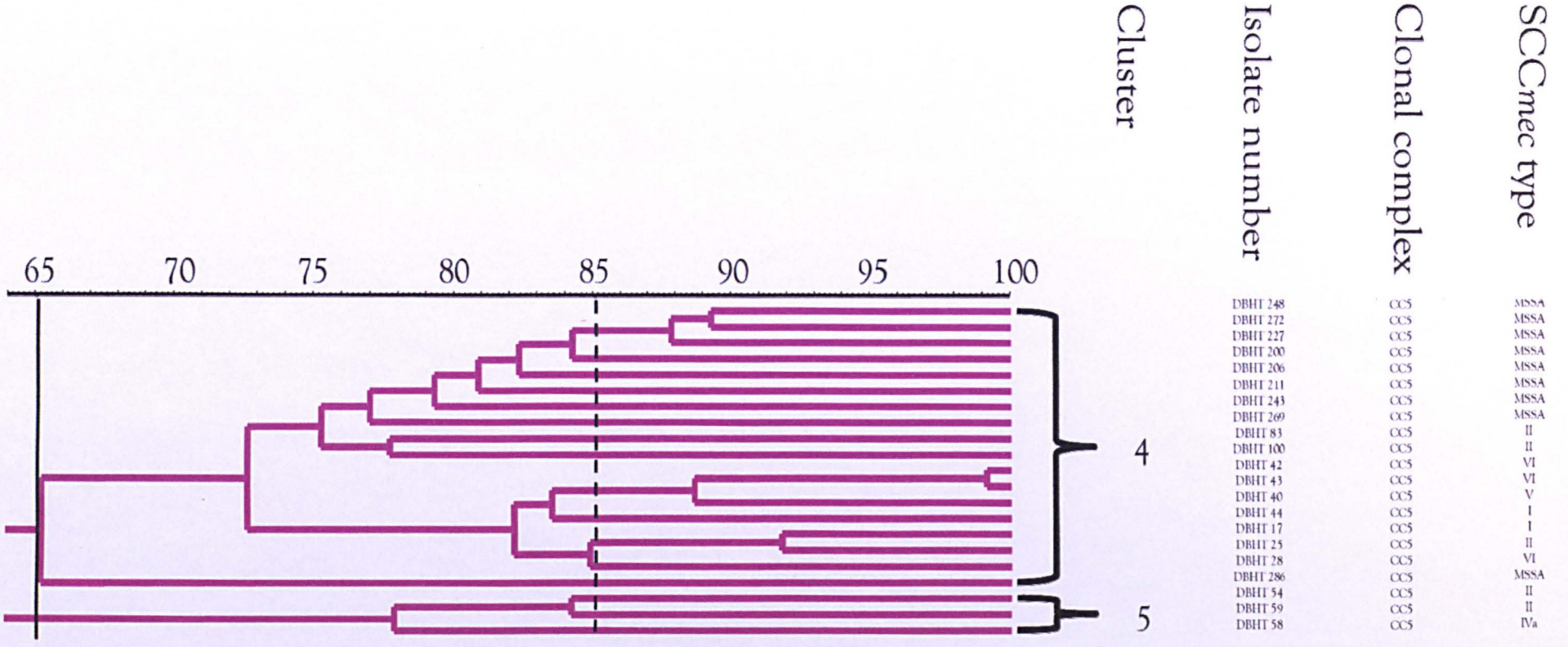
Figure 28 continued. Cluster 3 of UPGMA-derived dendrogram based on *Bgl*II+A and *Csp*6I+0 FAFLP data



DBHT 122	CC22	IVh
DBHT 114	CC22	IVh
DBHT 97	CC22	IVh
DBHT 82	CC22	IVh
DBHT 105	CC22	IVh
DBHT 190	CC22	IVh
DBHT 102	CC22	IVh
DBHT 108	CC22	IVh
DBHT 106	CC22	IVh
DBHT 156	CC22	IVa
DBHT 176	CC22	-
DBHT 148	CC22	IVa
DBHT 173	CC22	IVa
DBHT 119	CC22	II
DBHT 139	CC22	IVh
DBHT 164	CC22	IV
DBHT 95	CC22	IVh
DBHT 166	CC22	IVh
DBHT 68	CC22	IVh
DBHT 91	CC22	IVh
DBHT 298	CC22	V
DBHT 22	CC22	IVa
DBHT 24	CC22	IVd
DBHT 3	CC22	IVh
DBHT 7	CC22	IVh
DBHT 6	CC22	IVh
DBHT 8	CC22	IVh
DBHT 4	CC22	MSSA
DBHT 5	CC22	MSSA
DBHT 21	CC22	IVh
DBHT 32	CC22	IVh
DBHT 33	CC22	IVh
DBHT 34	CC22	IVh
DBHT 31	CC22	IVh
DBHT 23	CC22	IVh
DBHT 20	CC22	IV
DBHT 45	CC22	V
DBHT 18	CC22	II
DBHT 235	CC22	MSSA
DBHT 78	CC22	IV

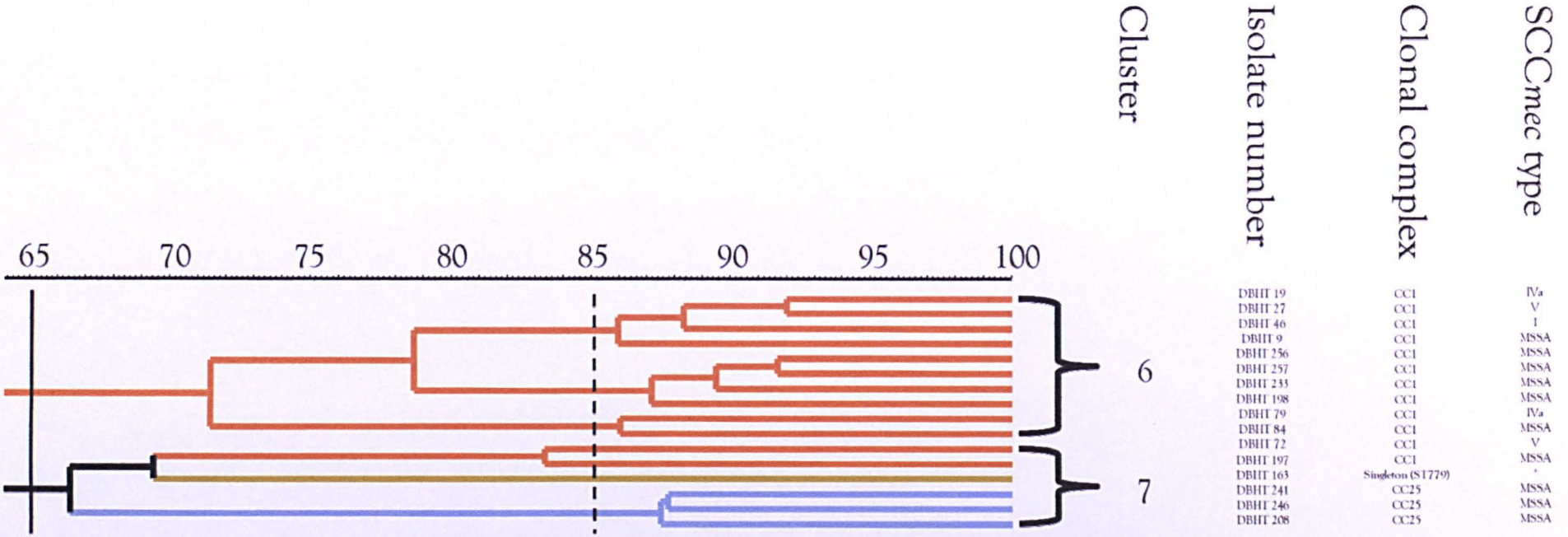
Cluster 3 was based on 65 % similarity cut-off (shown as vertical black line). The dashed line represents 85 % similarity cut-off. The clonal complex (CC) and SCC*mec* type of each isolate is shown on the right.

Figure 29. Clusters 4 and 5 of UPGMA-derived dendrogram based on *Bgl*II+A and *Csp*6I+0 FAFLP data



Clusters 4 and 5 were based on 65 % similarity cut-off (shown as vertical black line). The dashed line represents 85 % similarity cut-off. The clonal complex (CC) and SCC*mec* type of each isolate is shown on the right.

Figure 30. Clusters 6 and 7 of UPGMA-derived dendrogram based on *Bgl*II+A and *Csp*6I+0 FAFLP data



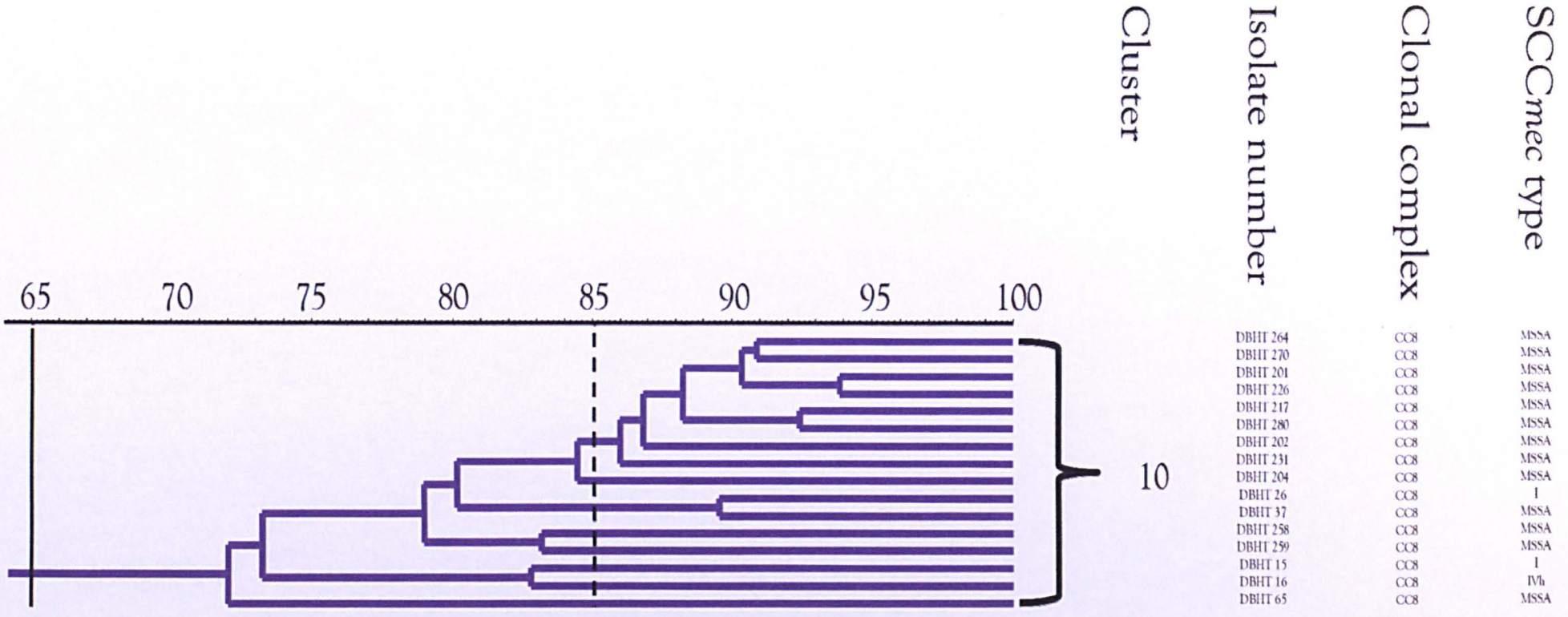
Clusters 6 and 7 were based on 65 % similarity cut-off (shown as vertical black line). The dashed line represents 85 % similarity cut-off. The clonal complex (CC) and SCC*mec* type of each isolate is shown on the right. The different colour lines represent a CC: a brown line represents CC1, a light blue line represents CC25 and a green line represents a singleton sequence type (ST779).

Cluster 8 consisted of two isolates which exhibited individual profiles which were 32 % divergent. Thirty-six AFs were common amongst the two isolates whilst 33 AFs were polymorphic. Both MSSA isolates were assigned to CC6 and CC101. Four isolates amongst cluster 9 presented individual profiles with up to 26.5 % difference. Thirty similar AFs were identified amongst the isolates and another 30 AFs were variable. Cluster 9 isolates belonged to CC7, one isolate belonged to SCC*mec* type IV whilst the remaining three isolates were MSSA.

Sixteen isolates constituting cluster 10 displayed distinct profiles and ≤28 % difference. Profiles shared 38 AFs whilst 87 AFs were different. Fifty-three percent (n=16) of all CC8 isolates constituted cluster 10 and 13 isolates were MSSA. The presence of SCC*mec* type I and IV was exhibited amongst three isolates (Figure 31).

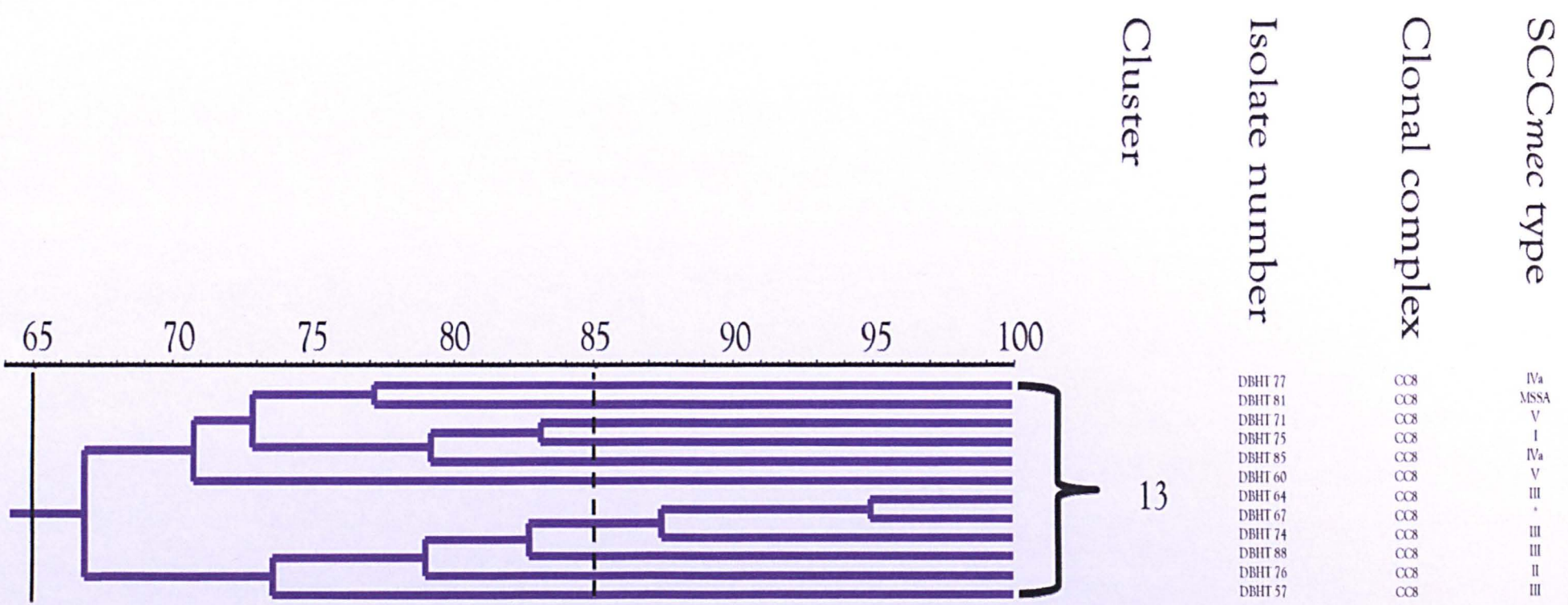
Each of the 12 isolates within cluster 13 displayed individual profiles and up to 33 % divergence. Isolates displayed 24 similar AFs and 97 variable AFs. Cluster 13 consisted of 40 % of CC8 isolates, 11 isolates belonged to one of six SCC*mec* types and one isolate was MSSA (Figure 32).

Figure 31. Cluster 10 of UPGMA-derived dendrogram based on *Bgl*II+A and *Csp*6I+0 FAFLP data



Cluster 10 was based on 65 % similarity cut-off (shown as vertical black line). The dashed line represents 85 % similarity cut-off. The clonal complex (CC) and SCC*mec* type of each isolate is shown on the right.

Figure 32. Cluster 13 of UPGMA-derived dendrogram based on *Bgl*II+A and *Csp*6I+0 FAFLP data

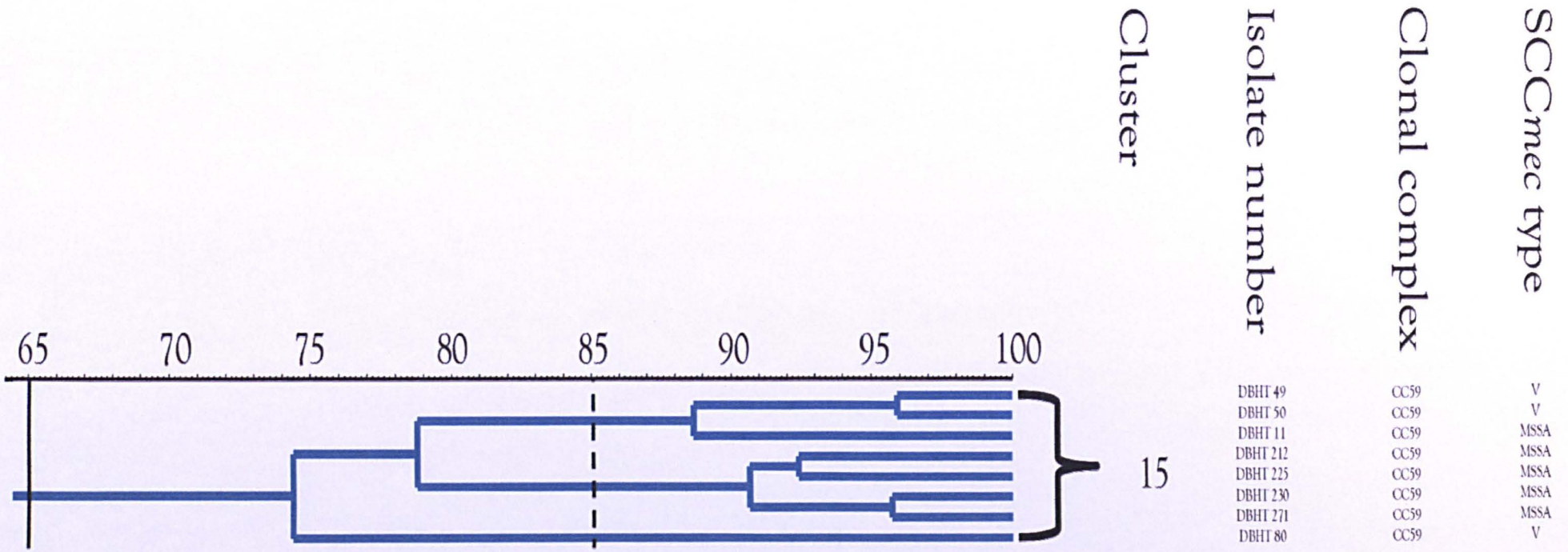


Cluster 13 was based on 65 % similarity cut-off (shown as vertical black line). The dashed line represents 85 % similarity cut-off. The clonal complex (CC) and SCC*mec* type of each isolate is shown on the right.

Two isolates belonging to cluster 11 exhibited indistinguishable profiles and displayed 50 similar AFs. Isolates were assigned to CC672 and were assigned to SCC*mec* type IVa. Cluster 12 was composed of three CC80 isolates which displayed unique profiles and up to 33.5 % divergence. These profiles displayed 39 conserved AFs and 40 polymorphic AFs. Cluster 14 grouped two isolates which exhibited individual profiles and 35 % difference. Forty-six AFs were common amongst the two isolates whilst 37 AFs were variable. Both isolates were assigned to ST93 and belonged to SCC*mec* type IVa. Cluster 15 had eight isolates which displayed distinct profiles and between 4 and 25.5 % divergence. Thirty-nine similar AFs and 47 different AFs were identified amongst profiles. Cluster 15 isolates constituted all CC59 isolates in this study; three isolates exhibited SCC*mec* type V and the remaining five isolates were MSSA (Figure 33).

Cluster 16 comprised of 35 isolates, each of which displayed individual profiles. Between 3.5 to 30.5 % divergence was exhibited amongst the isolates. Profiles shared 31 AFs and displayed 135 variable AFs. Ninety-seven percent of CC30 isolates constituted cluster 16; 12 were assigned to SCC*mec* type II or IV and 23 were MSSA (Figure 34). Three isolates amongst cluster 17 displayed unique profiles and up to 31.5 % divergence. Isolates displayed 38 common AFs and 37 different AFs. Of the three CC97 isolates of cluster 17, two isolates belonged to SCC*mec* type IV and V and one isolate was MSSA.

Figure 33. Cluster 15 of UPGMA-derived dendrogram based on *Bgl*II+A and *Csp*6I+0 FAFLP data



Cluster 15 was based on 65 % similarity cut-off (shown as vertical black line). The dashed line represents 85 % similarity cut-off. The clonal complex (CC) and SCC*mec* type of each isolate is shown on the right.

Figure 34. Cluster 16 of UPGMA-derived dendrogram based on *Bgl*II+A and *Csp*6I+0 FAFLP data

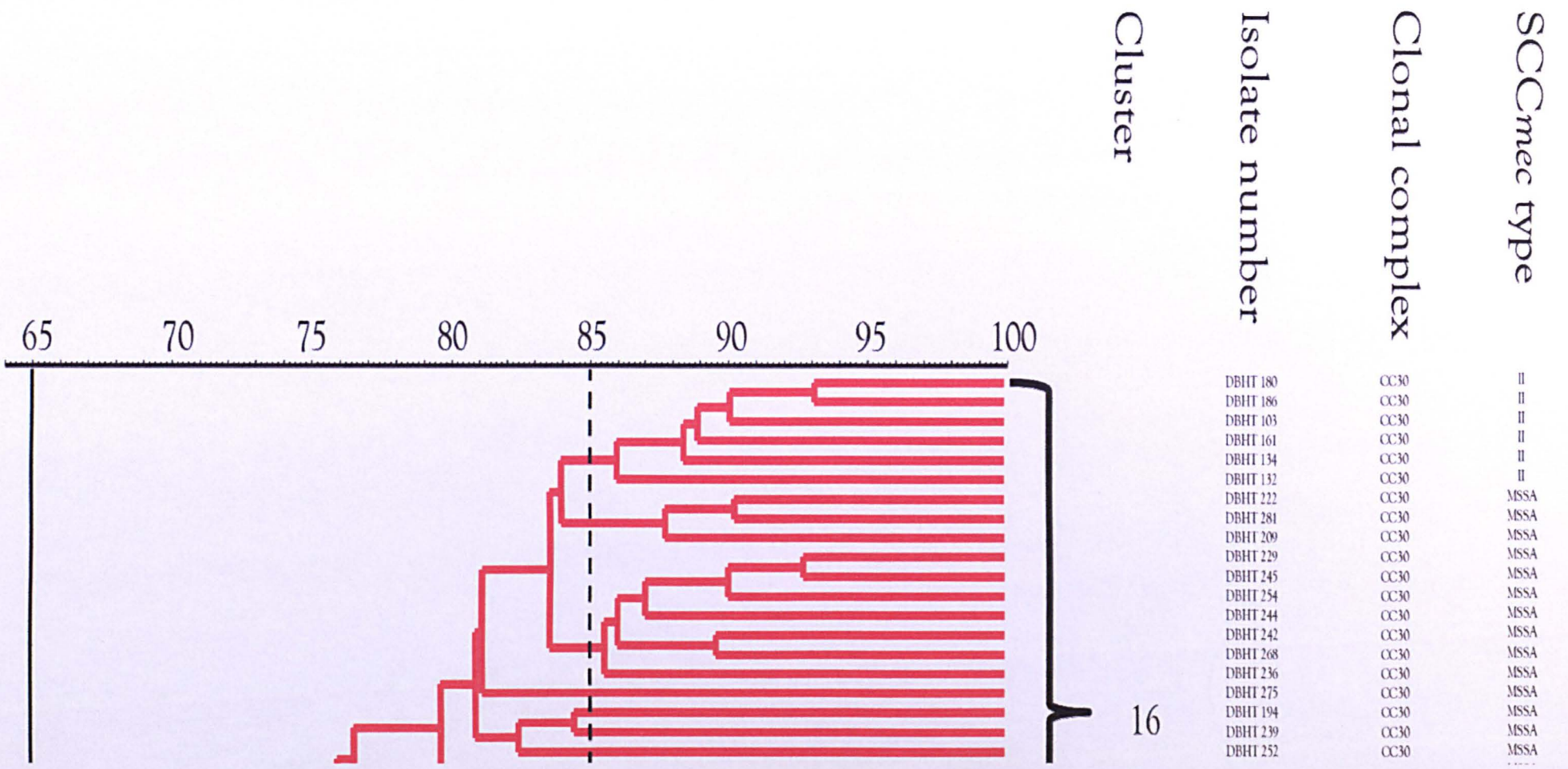
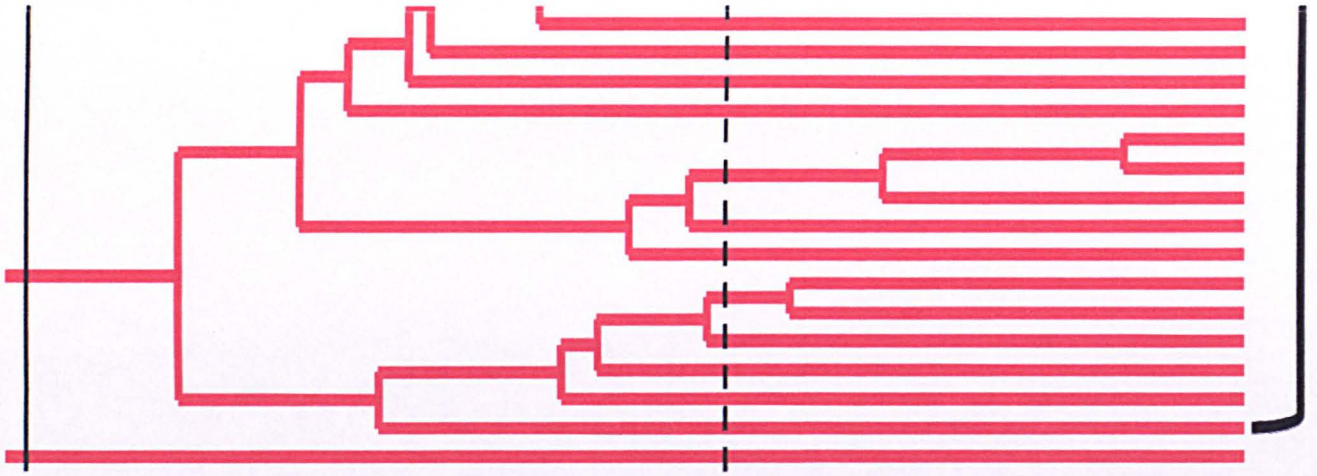


Figure 34 continued. Cluster 16 of UPGMA-derived dendrogram based on *Bgl*II+A and *Csp*6I+0 FAFLP data



DBHT 263	CC30	MSSA
DBHT 203	CC30	MSSA
DBHT 219	CC30	MSSA
DBHT 98	CC30	II
DBHT 61	CC30	II
DBHT 62	CC30	II
DBHT 55	CC30	II
DBHT 51	CC30	II
DBHT 66	CC30	MSSA
DBHT 36	CC30	MSSA
DBHT 38	CC30	MSSA
DBHT 35	CC30	MSSA
DBHT 30	CC30	IVc
DBHT 10	CC30	MSSA
DBHT 291	CC30	MSSA
DBHT 287	CC30	MSSA

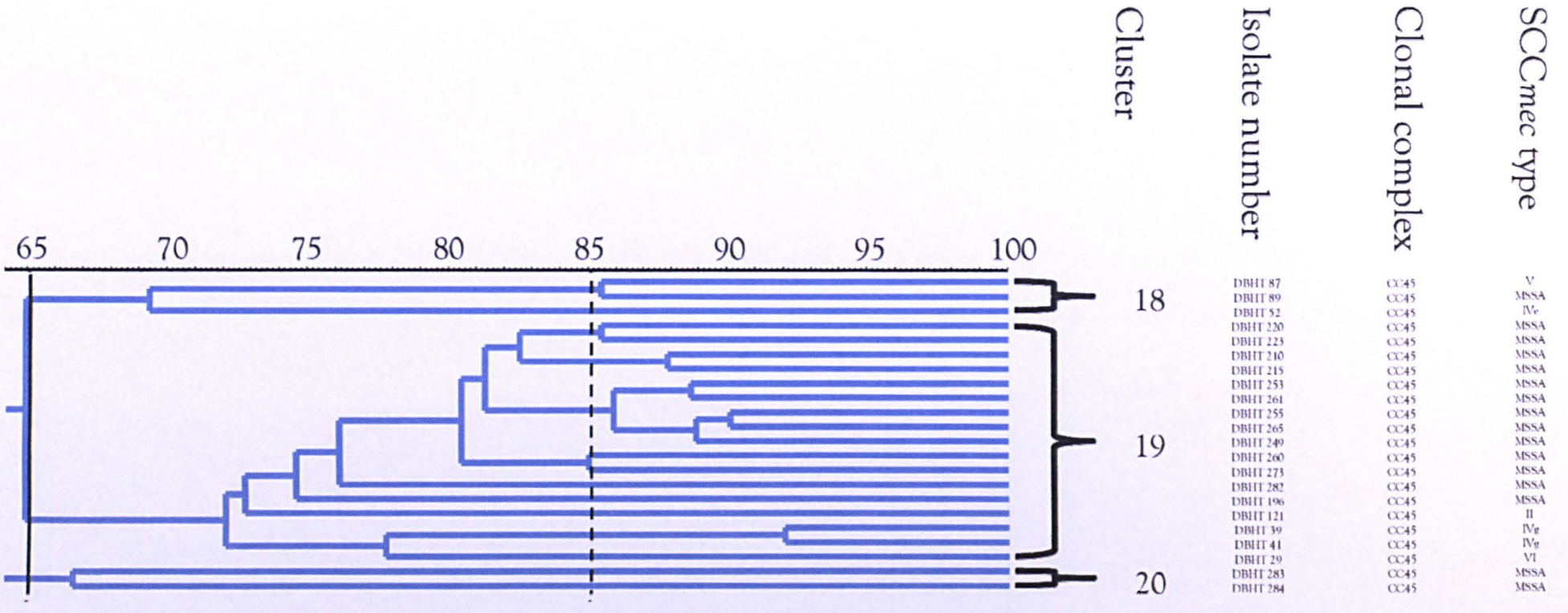
Cluster 16 was based on 65 % similarity cut-off (shown as vertical black line). The dashed line represents 85 % similarity cut-off. The clonal complex (CC) and SCC*mec* type of each isolate is shown on the right.

Cluster 18 consisted of three isolates displaying distinct profiles and up to 30.5 % difference. Thirty-seven AFs were conserved and 40 variable AFs were identified amongst profiles. The isolates constituted 14 % of CC45; one isolate each belonged to SCC*mec* type IV, V or was MSSA. The 17 isolates within cluster 19 exhibited individual profiles and up to 28 % divergence. Profiles displayed 37 similar AFs and 88 polymorphic AFs. Cluster 19 was composed of 77 % of CC45 isolates; four CC45 isolates exhibited one of three SCC*mec* types and 13 were MSSA. Both isolates within cluster 20 displayed individual profiles and 38 % divergence. Profiles shared 59 common AFs and 57 different AFs. The two MSSA isolates constituted 9 % of CC45 isolates (Figure 35).

Fourteen isolates amongst cluster 21 displayed individual profiles and 33 % divergence. Thirty-two conserved AFs and 62 variable AFs were identified amongst profiles. Isolates were assigned to CC9, 15, 20 and 121 and were MSSA. One CC15 isolate displayed >35 % divergence (Figure 36).

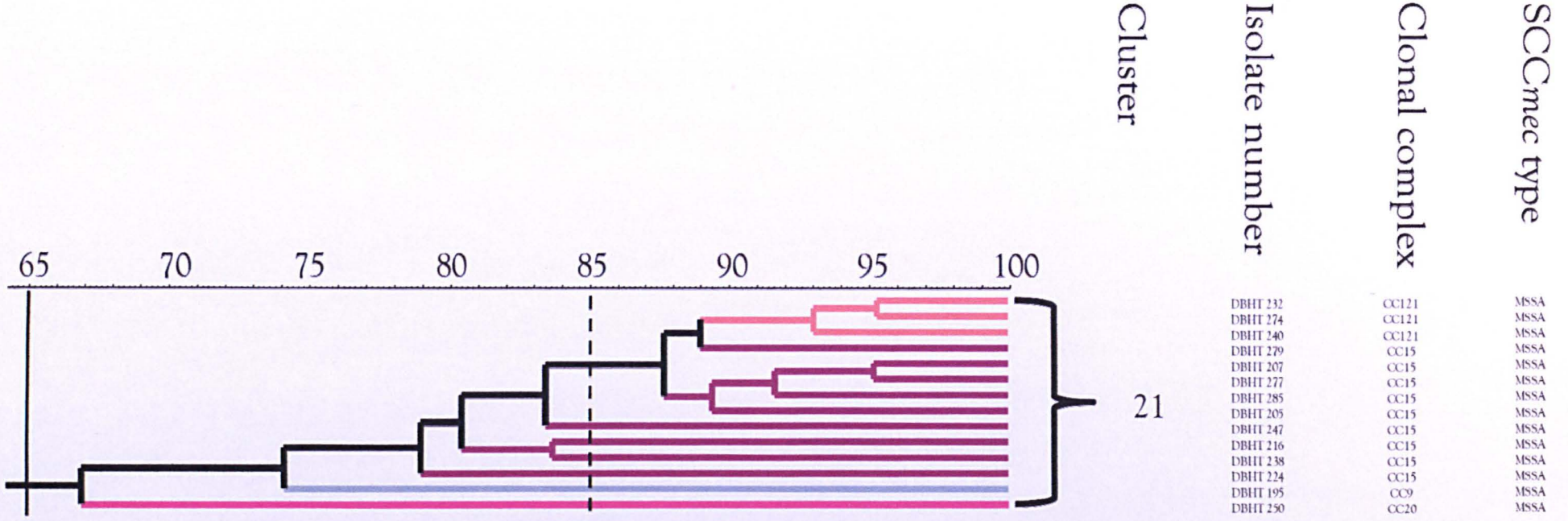
The majority of the 21 clusters exhibited isolates from a single CC. Clusters 2 and 3 harboured seven and 91 % of CC22 isolates. Isolates constituting cluster 2 displayed up to 40.5 % divergence from the 117 isolates of cluster 3. CC5 isolates were grouped within cluster 4 (78 %) and 5 (13 %) and exhibited up to 37.5 % divergence. Eighty-three percent and 17 % of CC1 isolates were assigned to clusters 6 and 7 respectively, and they differed by up to 42.5 %. Clusters 10 and 13 harboured 53 % and 40 % of CC8 isolates respectively, and isolates displayed up to 52.5 % difference between the two clusters. CC45 isolates belonging to clusters 18 (14 %), 19 (77 %) and 20 (9 %) displayed up to 38 % divergence. The fourteen divergent profiles were assigned to seven CCs (CC5, 8, 12, 15, 22, 30 and 88) and three singleton STs.

Figure 35. Clusters 18, 19 and 20 of UPGMA-derived dendrogram based on *Bgl*II+A and *Csp*6I+0 FAFLP data



Clusters 18, 19 and 20 were based on 65 % similarity cut-off (shown as vertical black line). The dashed line represents 85 % similarity cut-off. The clonal complex (CC) and SCC*mec* type of each isolate is shown on the right.

Figure 36. Cluster 21 of UPGMA-derived dendrogram based on *Bgl*II+A and *Csp*6I+0 FAFLP data



Cluster 21 was based on 65 % similarity cut-off (shown as vertical black line). The dashed line represents 85 % similarity cut-off. The clonal complex (CC) and SCC*mec* type of each isolate is shown on the right. The different colour lines represent a CC: a grey line represents CC9, a purple line represents CC15, a dark pink line represents CC20 and a light pink line represents CC121.

To reflect the FAFLP-based divergence between the isolates, a ≥85 % similarity cut-off was further applied to the dendrogram. Five sub-clusters were identified amongst the isolates in clusters 2 and 3. Zero to 32 AF differences were observed within each sub-cluster; conversely five to 94 polymorphic AFs were observed between sub-clusters. Each sub-cluster consisted of CC22 isolates; two sub-clusters harboured MRSA isolates whilst three harboured both MRSA and MSSA isolates. Three sub-clusters were identified amongst CC5 isolates of cluster 4. Up to 16 variable AFs were identified within each sub-cluster whilst up to 28 polymorphic AFs were observed between sub-clusters. One sub-cluster was composed of MSSA isolates and the remaining two harboured MRSA isolates. The CC1 isolates within cluster 6 exhibited three sub-clusters. Up to seven AF differences were observed within each sub-cluster and up to 28 polymorphic AFs were identified between sub-clusters. MSSA isolates constituted one sub-cluster and two sub-clusters were composed of MRSA and MSSA. One sub-cluster each consisting of CC7 and CC25 MSSA were identified amongst cluster 9 and cluster 7 respectively.

Three sub-clusters were displayed amongst cluster 10 and 13 isolates. Up to 15 AF differences were identified within sub-clusters whilst up to 35 polymorphic AFs were observed between sub-clusters. One sub-cluster was composed of MRSA, one of MSSA, and one of MRSA and MSSA. CC59 isolates within cluster 15 exhibited two sub-clusters. Up to nine variable AFs were identified within sub-clusters whilst up to 21 polymorphic AFs were observed between sub-clusters. MSSA isolates constituted one sub-cluster whilst the second was composed of MRSA and MSSA. CC30 isolates within cluster 16 exhibited five sub-clusters. Up to 19 polymorphic AFs were displayed within sub-clusters and up to 39 AF differences were observed between sub-clusters. Two sub-clusters were composed of MRSA isolates which belonged to SCC*mec* type II, and all isolates within these sub-clusters

excluding one were assigned to EMRSA-16. The three remaining sub-clusters were formed of MSSA isolates. Six sub-clusters were exhibited amongst CC45 isolates of cluster 18 and 19. Up to 18 polymorphic AFs were observed within each sub-cluster; conversely up to 47 variable AFs were observed between sub-clusters. Four sub-clusters harboured MSSA, one MRSA, and one MRSA and MSSA. One sub-cluster was identified within cluster 21 which harboured eight MSSA isolates assigned to CC15 and CC121.

3.3.2.5. Correlation of FAFLP data within isolate collections

All isolates within clusters 1, 8, 11, 18, 20 or 21 belonged to the same isolate collection. The remaining fifteen clusters consisted of isolates obtained from two or more collections. Isolates within each of the identified sub-clusters belonged to the same isolate collection. Cluster 4 isolates (n=18) belonging to CC5 were obtained from three collections whilst those belonging to cluster 5 (n=3) were obtained from the Glasgow or Reference collection. The CC8 isolates in cluster 10 (n=16) belonged to the SRL, Glasgow and BSAC collections. However, 11 CC8 isolates within cluster 13 (92 %) were from the Glasgow collection. Of the cluster 2 isolates assigned to CC22 (n=9), seven (78 %) were obtained from CMPHL collection and the remaining two (22 %) from BSAC collection. On the other contrary, cluster 3 isolates belonging to CC22 (n=117) consisted of isolates from four collections. Cluster 18 isolates belonging to CC45 (n=3) were from the Glasgow collection. Cluster 19 isolates (n=17) were from SRL and BSAC collections and cluster 20 isolates (n=2) were from BSAC collection. Clusters harbouring CC1, 30 and 59 isolates were assigned to three or four collections.

Eight of the 50 isolates from SRL were from a single patient. Six (cluster 3) of these were CC22 and exhibited ≥98 % similarity whilst the remaining two were CC672 isolates (cluster 11) and exhibited indistinguishable profiles. Multiple isolates within SRL collection which were obtained from the same hospital or ward were assigned to the same lineage and exhibited ≥84.5 % similarity. The reference strains clustered with other isolates of the same lineage. CC8 and 45 isolates from the Glasgow collection grouped in clusters 13 and 18; the CC8 isolates exhibited up to 52.5 % difference from the remaining CC8 isolates in the study. Profiles from the BSAC collection displayed no correlation with the centre of isolation. A majority of the CMPHL isolates (n=7; 70 %) were assigned to cluster 2; ≥65.5 % similarity was exhibited amongst these isolates.

3.3.2.6. Discriminatory power and confidence interval values

The D and CI values for FAFLP data were calculated for all isolates. Both values were also calculated for isolates from each of the five collections and for the eight major CCs: CC1, 5, 8, 15, 22, 30, 45 and 59. The D value for all FAFLP data was 0.999 with a CI of 0.999 to 1.000. The calculated D values for isolates from the same collection was 1.0 in each case except for those from SRL (D=0.997) and BSAC (D=0.999) collections. The calculated D values for isolates from the eight CCs were 1.0 except for those from CC22 (D=0.999).

3.3.2.7. Correlation of FAFLP amplified fragments (AFs) with lineages

Initially FAFLP profiles within isolates assigned to one of seven major lineages (CC1, 5, 8, 22, 30, 45 and 59) were compared in Stata software. Of the 500 polymorphic AFs amongst

all isolates, 73 fragments were exclusive to isolates assigned to one of the seven lineages. One AF was specific to CC1, nine AFs to CC5, 28 AFs to CC8, 24 AFs to CC22, six AFs to CC30, three AFs to CC45 and two AFs to CC59. Seventy-two of these AFs were present in up to three isolates of a specific lineage and the one remaining AF was present in 90 % of CC22 isolates (data not shown). Therefore although these AFs were lineage-specific, they were not present in all or the majority of isolates of a particular lineage. Fragments which were common, polymorphic and specific to isolates of each lineage were subsequently identified manually (Table 27). Relative to the number of isolates assigned to each CC, the largest number of common AFs was demonstrated within CC1 followed by CC59 and isolates within CC5 and 8 displayed the maximum numbers of polymorphic AFs. A total of 50 lineage-specific AFs were identified. The largest numbers of lineage-specific AFs were exhibited amongst CC30, 45 and 59 isolates.

Table 27. Analysis of FAFLP profiles with primers *Bgl*I+A and *Csp*6I+0

CC	CC1	CC5	CC8	CC22	CC30	CC45	CC59	Total
Common AFs	28	31	23	34	31	29	36	212
Lineage-specific AFs	1	4	2	6	13	11	13	50
Polymorphic AFs	69	181	209	292	165	145	50	1111

Number of amplified fragments (AFs) common to all isolates of one of seven clonal complexes (CCs), polymorphic amongst isolates of a particular CC and specific to a particular CC are shown.

3.3.2.8. Lineage-specific AFs

Fifty lineage-specific AFs were identified and the sequences of 15 of these AFs were determined experimentally from isolates representative of one of seven CCs: CC1, 5, 8, 22, 30, 45 and 59. Of the 50 AFs, the remaining 35 AFs comigrated on agarose gels with AFs of a similar size and were hence difficult to sequence. Alignments based on BLAST search results (from up to 27 WGSs) of the lineage-specific AF sequences were performed. These AFs were found for CC30 (n=8), CC45 (n=2), CC59 (n=2), CC5 (n=1), CC8 (n=1) and CC22 (n=1) (Table 28). PCR primers designed for these 15 AFs were evaluated on 48 isolates. Sequences of each AF were obtained in seven or more CCs. Therefore although the size of these AFs were lineage-specific, similar sequences were identified in majority of the lineages. This indicated these sequences harboured lineage-specific variations, therefore these sequences were compared between lineages.

Of the remaining 35 lineage-specific AFs, sequences for 34 were obtained via *in silico* analysis of WGSs; one AF size was not detected *in silico* and was not investigated further. Of the 34 AFs, 11 AFs were specific to CC59, eight to CC45, five each to CC22 and 30, three to CC5 and one each to CC1 and 8. Six of the 34 AF sequences were detected in up to four CCs. The remaining 28 sequences were detected in isolates assigned to six to 10 different CCs (Table 29). Although the sizes of these AFs were lineage-specific, experimental data obtained from determining the sequence of these AFs showed the presence of these sequences in the majority of other lineages. This indicated that the AF sequences also harboured lineage-specific variations which was analysed further.

Table 28. Sequence determination results of lineage-specific AFs

Lineage-specific AF	Size of AF (bp)*	Lineage specificity*	Presence in lineage†
F1	286	CC5	CC1, 5, 15, 22, 30, 45 and 97
F2	289	CC8	CC1, 5, 8, 15, 22, 30, 45, 59 and 97
F3	141	CC22	CC1, 5, 8, 15, 22, 30, 45, 59, 97 and 672
F4	97	CC30	CC1, 5, 8, 15, 22, 30, 45, 59 and 97
F5	156	CC30	CC1, 5, 15, 22, 30, 59, 97 and 672
F6	182	CC30	CC1, 5, 8, 15, 22, 30, 45, 59, 97 and 672
F7	277	CC30	CC1, 5, 8, 15, 22, 30, 45, 59, 97 and 672
F8	307	CC30	CC1, 5, 8, 15, 22, 30, 45, 59 and 97
F9	389	CC30	CC1, 5, 8, 15, 22, 30, 45, 59 and 97
F10	479	CC30	CC1, 5, 8, 15, 22, 30, 45, 59 and 97
F11	528	CC30	CC1, 5, 8, 15, 22, 30, 45, 59 and 97
F12	82	CC45	CC1, 5, 8, 15, 22, 30, 45, 59, 97 and 672
F13	137	CC45	CC1, 5, 8, 15, 22, 30, 45, 59, 97 and 672
F14	85	CC59	CC1, 5, 8, 15, 22, 30, 45, 59 and 97
F15	340	CC59	CC1, 5, 8, 15, 22, 30, 45, 59 and 97

*Sizes of amplified fragments (AFs), generated using FAFLP analysis with primers *Bgl*II+A and *Csp*6I+0, which were identified in isolates of a particular lineage.

†Lineages within which sequences of lineage-specific AFs were identified.

3.3.2.9. Sequence variation amongst lineage-specific AFs

Sequence variation was observed amongst fragments of a precise size specific to a particular lineage. Sequence variation was observed amongst 39 of the 50 lineage-specific AFs. Of these AFs, 12 were sequences obtained experimentally whilst 27 were obtained via *in silico* analysis.

Table 29. Lineage-specific AF sequences obtained via *in silico* analysis

Lineage-specific AF*	Size of AF (bp)*	Lineage specificity*	Loci of AF†	Accession number‡	AF present in lineages§
F16	355	CC1	60512-60839	NC_003923	CC1, 5, 8, 22, 30, 93 and 398
F17	368	CC5	2352726-2353073	NC_009782	CC1, 5, 8, 22, 30, 59, 93, 151, 398 and 425
F18	388	CC5	2353068-2353431	NC_009782	CC1, 5, 8, 22, 30, 59, 93, 151, 398 and 425
F19	414	CC5	485187-485577	NC_009782	CC1, 5, 8, 22, 30, 59, 93, 151, 398 and 425
F20	353	CC8	2293146-2293474	NC_007795	CC1, 5, 8, 22, 30, 59, 93, 151, 398 and 425
F21	107	CC22	2682498-2682584	HE681097	CC22
F22	117	CC22	432061-432157	HE681097	CC5, 8, 22, 93, 151 and 398
F23	270	CC22	2440502-2440749	HE681097	CC22, 30, 59 and 425
F24	331	CC22	2306520-2306830	HE681097	CC1, 5, 8, 22, 30, 59, 93, 151, 398 and 425
F25	411	CC22	1555575-1555960	HE681097	CC1, 8, 22, 30, 93, 105 and 398
F26	124	CC30	2703901-2704002	NC_002952	CC1, 5, 8, 22, 30, 59, 93, 151, 398 and 425
F27	247	CC30	2765286-2765506	NC_002952	CC1, 5, 8, 22, 30, 59, 93, 151, 398 and 425
F28	251	CC30	1314761-1314890	NC_002952	CC1, 5, 8, 22, 30, 59, 93, 151, 398 and 425
F29	295	CC30	92666-92936	NC_002952	CC30 and 239
F30	296	CC30	1373018-1373289	NC_002952	CC1, 5, 8, 22, 30, 59, 93, 151, 398 and 425
F31	102	CC45	1144078-1144159	NC_017349	CC1, 5, 8, 22, 30, 59, 93, 151, 398 and 425
F32	299	CC45	45394-45668	NC_002951	CC1, 5, 8 and 30
F33	317	CC45	1166118-1166412	NC_009782	CC1, 5, 8, 22, 30, 59, 93, 151, 398 and 425
F34	347	CC45	2326879-2327203	NC_007795	CC1, 5, 8, 22, 30, 59, 93, 151, 398 and 425
F35	354	CC45	2611039-2611372	NC_017338	CC1, 5, 8, 22, 30, 59, 93, 151, 398 and 425
F36	363	CC45	347212-347552	NC_009641	CC8 and 425

Table 29 continued. Lineage-specific AF sequences obtained via *in silico* analysis

Lineage-specific AF*	Size of AF (bp)*	Lineage specificity*	Loci of AF†	Accession number‡	AF present in lineages§
F37	496	CC45	N/A‖	N/A	N/A
F38	573	CC45	1105871-1106421	CP003166	CC1, 5, 8, 22, 30, 59, 93, 151, 398 and 425
F39	585	CC45	718453-719016	NC_009632	CC5, 8 and 93
F40	212	CC59	1887952-1888142	CP003166	CC1, 5, 8, 22, 30, 59, 93, 151, 398 and 425
F41	222	CC59	701113-701312	CP003166	CC1, 5, 8, 22, 30, 59, 93, 151, 398 and 425
F42	304	CC59	1554736-1555018	CP003166	CC1, 5, 8, 22, 30, 59 and 93
F43	350	CC59	2277176-2277500	CP003166	CC1, 5, 8, 22, 30, 59, 93, 151, 398 and 425
F44	389	CC59	290065-290430	CP003166	CC1, 5, 8, 22, 30, 59, 93, 151, 398 and 425
F45	429	CC59	1663979-1664389	CP003166	CC1, 5, 8, 22, 30, 59, 93, 151, 398 and 425
F46	447	CC59	2245902-2246327	CP003166	CC1, 5, 8, 22, 30, 59, 93, 151, 398 and 425
F47	537	CC59	2424067-2424581	CP003166	CC1, 5, 8, 22, 30, 59, 93, 151, 398 and 425
F48	551	CC59	447120-447646	CP003166	CC1, 5, 8, 22, 30, 59, 93, 151, 398 and 425
F49	571	CC59	1105871-1106421	CP003166	CC1, 5, 8, 22, 30, 59, 93, 151, 398 and 425
F50	599	CC59	1835-2410	CP003166	CC1, 5, 8, 22, 30, 59, 93, 151, 398 and 425

*Sizes of amplified fragments (AFs) generated using FAFLP analysis with primers *Bgl*I+A and *Csp*6I+0 which were identified in isolates of a particular lineage.

†Locus of AF based on whole-genome sequences (WGSs).

‡ NCBI database accession number of WGSs used for analysis.

§Lineages within which sequences of lineage-specific AFs were identified based on WGSs.

‖N/A, not applicable as lineage-specific AF was not detected *in silico*.

The sequences of six (F21, 23, 29, 32, 36 and 39) of the 50 AFs were amplified in four CCs or less, therefore sequence comparisons of these AFs could not be performed between lineages (Table 29). BLAST search results of these six AF sequences detected in four CCs revealed these partially encoded (did not encompass an entire ORF from a start to stop codon) SCC*mec* proteins, a transposase, a bacteriophage tape measure protein, an ABC transporter permease protein and a protein involved in the biosynthesis of cobalamin (vitamin B12). Four AFs (F4, 11, 14 and 25) displayed no sequence variation; three of these were obtained experimentally whilst one was obtained by *in silico* analysis. BLAST search results of these four AF sequences revealed that these partly encoded an UDP-N-acetylmuramoylalanine-D-glutamate ligase, transposase within the SCC*mec*, zinc-binding dehydrogenase and a bacteriophage DNA-directed DNA polymerase. One AF sequence (F37) could not be obtained experimentally or through *in silico* analysis as no AFs within ±7 bp of the expected AF size were detected (Table 28).

The sequence variations amongst the 39 AFs included SNPs, point mutations and indels specific to isolates of one lineage (Table 30). The point mutations included nine 1 bp insertions and 163 deletions. Four insertions between three to 24 bp and 25 deletions between three to 67 bp were also identified. All four insertions >1 bp and eight of the 25 deletions were multiples of three indicating the ORF is restored after the mutation. A total of 1204 SNPs, point mutations and indels were identified amongst the 39 AFs. Of these, 1003 were SNPs and 201 were insertions or deletions of 1 bp or greater. Between one and 111 variations were identified in each of the AF sequences. The maximum number of variations or SNPs were exhibited amongst the CC59-specific AF sized 551 bp (F48, n=111) whilst the least variations were identified in CC30- and CC45-specific AFs of 277 bp (F7, n=1) and 137 bp (F13, n=1). The CC8-specific 353 bp AF (F20) exhibited the greatest

number of indels (n=34). Twenty AF sequences exhibited no indels larger than 1 bp. The mean number of variations amongst all AFs specific to one lineage was identified. The maximum number of variations was identified amongst AFs specific to CC8 (n=188) whilst the minimum number was obtained within CC45 (n=95).

Table 30. Sequence variation in lineage-specific AFs

Lineage-specific AF	Size of AF (bp) in CC*	No. SNPs† per AF	No. SNPs‡ within CC										Variations specific to no. of CCs§
			CC1	CC5	CC8	CC22	CC30	CC45	CC59	CC97	CC672	ST93	
F16	355 (CC1)	40 (14)	8 (2)	6	6	4 (2)	6 (1)	-	-	-	-	10 (9)	6
F1	286 (CC5)	16 (8)	3 (1)	10 (6)	-	-	2	1 (1)	-	-	-	-	4
F17	368 (CC5)	47 (13, 1)	13 (1)	5 (2)	5 (1)	9 (6, 1)	5 (1)	-	3	-	-	7 (2)	7
F18	388 (CC5)	92 (27, 2)	22 (5)	16 (10)	17 (5, 1)	8 (4)	9	-	12 (1)	-	-	8 (3)	7
F19	414 (CC5)	34 (1)	1	1	1	3 (1)	17	-	4	-	-	7	7
F2	289 (CC8)	96 (30, 1)	8	13 (7)	2 (1)	15 (9)	7 (1)	27 (10)	14 (2)	10 (1)	-	-	8
F20	353 (CC8)	92 (33, 1)	25 (5)	11 (7)	16 (6)	14 (14)	9	-	11 (1)	-	-	6 (1)	7
F22	117 (CC22)	63 (0)	-	18	22	18	-	-	-	-	-	5	4
F3	141 (CC22)	34 (2)	-	1	-	-	16	12 (1)	4 (1)	1	-	-	5
F24	331 (CC22)	54 (16, 1)	12	4	4	19 (14, 1)	6 (1)	-	5	-	-	4 (1)	7
F26	124 (CC30)	41 (3)	1	-	-	-	1	-	-	-	-	39 (3)	3
F5	156 (CC30)	8 (0)	1	-	-	-	-	-	-	5	2	-	3
F6	182 (CC30)	27 (0)	-	-	-	1	2	24	-	-	-	-	3
F27	247 (CC30)	8 (1)	-	4	-	2 (1)	-	-	1	-	-	1	4

Table 30 continued. Sequence variation in lineage-specific AFs

Lineage-specific AF	Size of AF (bp) in CC*	No. SNPs† per AF	No. SNPs‡ within CC										Variations specific to no. of CCs§
			CC1	CC5	CC8	CC22	CC30	CC45	CC59	CC97	CC672	ST93	
F28	251 (CC30)	2 (0)	-	-	-	-	-	-	1	-	-	1	2
F7	277 (CC30)	1 (0)	-	1	-	-	-	-	-	-	-	-	1
F30	296 (CC30)	23 (1)	2	6 (1)	2	-	-	-	4	-	-	9	5
F8	307 (CC30)	68 (17)	8 (1)	5 (2)	2 (2)	15 (7)	4 (1)	20 (3)	9 (1)	5	-	-	8
F9	389 (CC30)	28 (5)	-	14 (3)	-	-	-	7 (1)	-	7 (1)	-	-	3
F10	479 (CC30)	16 (2)	1	2	-	1	6 (1)	-	5	1 (1)	-	-	6
F12	82 (CC45)	18 (0)	-	-	-	-	2	4	-	12	-	-	3
F31	102 (CC45)	9 (0)	1	-	-	2	3	-	3	-	-	-	4
F13	137 (CC45)	1 (0)	-	-	1	-	-	-	-	-	-	-	1
F33	317 (CC45)	3 (0)	1	-	-	1	-	-	-	-	-	1	3
F34	347 (CC45)	22 (0)	4	4	1	1	4	-	2	-	-	6	7
F35	354 (CC45)	26 (0)	4	1	4	2	2	-	1	-	-	12	7
F38	573 (CC45)	16 (1)	1	-	-	2	3	-	5 (1)	-	-	5	5
F40	212 (CC59)	15 (0)	-	1	2	2	5	-	1	-	-	4	6
F41	222 (CC59)	12 (0)	-	-	1	7	3	-	1	-	-	-	4
F42	304 (CC59)	3 (0)	-	-	-	-	-	-	3	-	-	-	1
F15	340 (CC59)	12 (0)	2	-	1	3	2	-	3	-	-	1	6
F43	350 (CC59)	21 (0)	4	4	1	1	4	-	2	-	-	5	7
F44	389 (CC59)	6 (0)	1	1	-	1	-	-	2	-	-	1	5
F45	429 (CC59)	17 (0)	1	4	3	3	2	-	1	-	-	3	7

Table 30 continued. Sequence variation in lineage-specific AFs

Lineage-specific AF	Size of AF (bp) in CC*	No. SNPs† per AF	No. SNPs‡ within CC										Variations specific to no. of CCs§
			CC1	CC5	CC8	CC22	CC30	CC45	CC59	CC97	CC672	ST93	
F46	447 (CC59)	48 (12, 3)	10 (1)	7	4 (1)	14 (11)	6 (1)	-	2	-	-	5 (1)	7
F47	537 (CC59)	43 (0)	-	3	2	34	2	-	2	-	-	-	5
F48	551 (CC59)	111 (2, 3)	10	83 (3)	1 (1)	4 (1)	5	-	2	-	-	6	7
F49	571 (CC59)	19 (1)	1	-	-	2	6	-	5 (1)	-	-	5	5
F50	599 (CC59)	12 (0)	-	6	-	2	1	-	1	-	-	2	5
	No. of SNPs* per CC	1204	145 (16)	231 (41)	98 (18)	190 (72)	140 (7)	95 (16)	109 (8)	41 (3)	2 (0)	153 (20)	

*Size of amplified fragments (AFs) generated using FAFLP analysis with primers *Bgl*I+A and *Csp*6I+0 specific to one clonal complex (CC; shown in brackets).

†Number includes number of single-nucleotide polymorphisms (SNPs), point mutations and indels in each AF.

‡The number of SNPs, point mutations and indels in each AF specific to one of 10 CC are shown. Numbers in brackets indicate the number of point mutations (1 bp insertions or deletions) and/or >1 bp indels. Black numbers in brackets indicate deletions and red numbers indicate insertions.

§Number of lineages for which specific variations were identified in each AF.

Each of the sequences exhibited variations specific to one of eight CCs. All four CC5-specific AF sizes displayed variations specific to CC1, 5 and 30. Both AFs specific to CC8 isolates displayed variations for at least six CCs. Amongst those AFs specific to CC22, 30 or 45 isolates, variations that were specific to nine or 10 CCs were identified. All AFs specific to CC59 displayed variations specific to this lineage. Amongst the 39 sequences, the total number of variations specific to isolates of one lineage varied from two (CC672) to 231 (CC5). The majority of variations were specific to isolates assigned to CC5 (19 %) followed by CC22 (16 %) and ST93 (13 %). Isolates belonging to CC5 (19 %) also exhibited the greatest number of SNPs, followed by CC30 (13 %) and ST93 (13 %) isolates. The maximum number of insertions and deletions that were ≥1 bp were exhibited amongst CC22 (36 %), CC5 (20 %) and ST93 (10 %) isolates.

Table 31. Results of PCRs targeting lineage-specific AF sequences

Lineage-specific AF	Size of AF (bp)	No. of isolates assigned to a CC
F1	286	24
F2	289	45
F3	141	22
F5	156	7
F6	182	10
F7	277	3
F8	307	46
F9	389	11
F10	479	27
F12	82	13
F13	137	7

Assignment of a collection of 48 blind isolates to a single clonal complex (CC) based on lineage-specific variations amongst one of 11 amplified fragment (AF) sequences. The remaining isolates were identified to belong to a subset of lineages or failed to amplify an amplicon.

A blind study was performed on 48 isolates to evaluate the specificity of lineage-specific variations amongst 11 of the 39 AF sequences. Based on these results, the isolates were assigned to a CC; some isolates could not be assigned to a single CC (Table 31). This indicated that these variations were relatively stable amongst a specific lineage.

3.3.2.10. Proteins encoded on lineage-specific AFs

Partial ORFs (a proportion of the region between a start and stop codon) and the proteins encoded by each of the 39 AF sequences were identified using sequences from the NCBI database which were aligned using BioEdit software (Table 32). Three AF sequences partially encoded a transposase enzyme (F1, F5 and F16). One of these (F5) also encoded, in part, a sodium transport protein. Eleven AF sequences encoded proteins required for the biosynthesis of essential cell components. Eight AFs (F2, F8, F9, F17, F18, F20, F24 and F46) partly encoded acetolactate synthase, a protein required for the biosynthesis of amino acids. Of these eight AFs, six also partially encoded an RNA-directed DNA polymerase (reverse transcriptase). Fragments F38, F45 and F49 encoded three additional proteins that play a role in the biosynthesis of haem. Proteins relating to the repair of damaged DNA were encoded on two AF sequences (F3 and F12). Proteins associated with DNA replication and recombination, such as a single-stranded DNA-binding protein, were encoded on eight AFs (F2, F7, F8, F9, F17, F18, F20 and F50).

Functions of a further five AF sequences were related to antimicrobial resistance. Two of these AFs encoded a multidrug efflux system protein (F34 and F43) and one encoded a protein implicated in the metabolism of a herbicide (F15). F41 partially encoded a protein which conferred bacitracin resistance and F44 partly encoded a protein which altered their

antimicrobial tolerance. Bacteriophage proteins were encoded on two ORFs; one partly encoded a capsid morphogenesis protein (F30) whilst the other for a portal protein (F42). Two AFs (F10 and F13) harboured ORFs relating to proteins with homology to proteins involved in programmed cell death (apoptosis) in eukaryotes. Lipoproteins were encoded on three AF sequences (F19, 22 and 48); these proteins have a structural function or are enzymes or toxins. Four AF sequences (F26, F27, F28 and F47) encoded proteins which are essential for the respiration, growth or cell division of *S. aureus*. Of the remaining five fragment sequences, four (F6, F31, F35 and F40) encoded enzymes such as hydrolases, proteases and methyltransferases and one AF (F33) encoded a hypothetical protein.

Alignment of the 39 AF sequences and the translated amino acid sequences identified one to 95 SNPs amongst ORFs of 37 sequences (Table 32). The remaining AFs did not harbour SNPs within ORFs. One to 41 synonymous SNPs were identified in ORFs and one to 62 SNPs were non-synonymous. Eighteen (49 %) of the 37 AFs exhibited a greater proportion of synonymous SNPs as compared to non-synonymous SNPs amongst ORFs. Of the remaining 19 AFs, 18 displayed more non-synonymous SNPs than synonymous SNPs and one AF displayed an equal proportion of synonymous and non-synonymous SNPs.

Table 32. Proteins and single-nucleotide polymorphisms encoded on lineage-specific AF sequences

Lineage-specific AF	Size of AF (bp) in CC*	No. SNPs per AF (no. in ORFs)†	No. synonymous and non-synonymous SNPs in coding regions‡									No. of synonymous and non-synonymous SNPs§	ORFs within AF sequence¶
			CC1	CC5	CC8	CC22	CC30	CC45	CC59	CC97	ST93		
F16	355 (CC1)	40 (37)	1	1	1	0	1	0	0	0	1	5	Transposase
			4	5	5	4	5	0	0	0	9	32	
F1	286 (CC5)	16 (14)	0	2	0	0	1	1	0	0	0	4	Transposase
			2	7	0	0	1	0	0	0	0	10	
F17	368 (CC5)	47 (25)	2	0	1	0	0	0	1	0	0	4	Acetolactate synthase and RNA-directed DNA polymerase (Reverse transcriptase)
			11	1	3	0	3	0	1	0	2	21	
F18	388 (CC5)	92 (38)	0	1	1	1	1	0	4	0	0	8	Acetolactate synthase and RNA-directed DNA polymerase (Reverse transcriptase)
			8	4	6	4	3	0	2	0	3	30	
F19	414 (CC5)	34 (34)	0	1	1	3	6	0	2	0	4	17	*Staphylococcus* tandem lipoprotein
			1	0	0	0	11	0	2	0	3	17	
F2	289 (CC8)	96 (73)	0	4	0	0	0	0	3	4	0	8	Acetolactate synthase and RNA-directed DNA polymerase (Reverse transcriptase)
			7	7	2	13	5	20	3	5	0	62	

Table 32 continued. Proteins and single-nucleotide polymorphisms encoded on lineage-specific AF sequences

Lineage-specific AF	Size of AF (bp) in CC*	No. SNPs per AF (no. in ORFs)†	No. synonymous and non-synonymous SNPs in coding regions‡									No. of synonymous and non-synonymous SNPs§	ORFs within AF sequence¶
			CC1	CC5	CC8	CC22	CC30	CC45	CC59	CC97	ST93		
F20	353 (CC8)	92 (52)	1	0	2	4	1	0	4	0	0	12	Acetolactate synthase and RNA-directed DNA polymerase (Reverse transcriptase)
			11	8	4	8	4	0	2	0	3	40	
F22	117 (CC22)	63 (63)	0	8	6	4	0	0	0	0	2	20	*Staphylococcus* tandem lipoprotein
			0	10	16	14	0	0	0	0	3	43	
F3	141 (CC22)	34 (0)	0	0	0	0	0	0	0	0	0	0	Recombination protein RecR
			0	0	0	0	0	0	0	0	0	0	
F24	331 (CC22)	54 (45)	1	1	0	3	1	0	2	0	1	9	Acetolactate synthase
			11	1	3	14	3	0	2	0	2	36	
F26	124 (CC30)	41 (2)	1	0	0	0	1	0	0	0	0	2	Pyruvate oxidase or thiamine pyrophosphate-requiring enzyme
			0	0	0	0	0	0	0	0	0	0	
F5	156 (CC30)	8 (1)	0	0	0	0	0	0	0	0	0	0	Sodium transport protein and transposase
			0	0	0	0	0	0	0	1	0	1	

Table 32 continued. Proteins and single-nucleotide polymorphisms encoded on lineage-specific AF sequences

Lineage-specific AF	Size of AF (bp) in CC*	No. SNPs per AF (no. in ORFs)†	No. synonymous and non-synonymous SNPs in coding regions‡									No. of synonymous and non-synonymous SNPs§	ORFs within AF sequence¶
			CC1	CC5	CC8	CC22	CC30	CC45	CC59	CC97	ST93		
F6	182 (CC30)	27 (27)	0	0	0	0	1	20	0	0	0	21	Zinc-dependent metalloprotease
			0	0	0	1	1	4	0	0	0	6	
F27	247 (CC30)	8 (3)	0	1	0	1	0	0	0	0	0	2	L-lactate dehydrogenase 2
			0	1	0	0	0	0	0	0	0	1	
F28	251 (CC30)	2 (2)	0	0	0	0	0	0	1	0	1	2	DNA translocase FtsK (cell division protein)
			0	0	0	0	0	0	0	0	0	0	
F7	277 (CC30)	1 (0)	0	0	0	0	0	0	0	0	0	0	Single-stranded DNA-binding protein
			0	0	0	0	0	0	0	0	0	0	
F30	296 (CC30)	23 (22)	1	0	1	0	0	0	2	0	2	6	Bacteriophage putative capsid morphogenesis protein
			1	5	1	0	0	0	2	0	7	16	
F8	307 (CC30)	68 (49)	0	1	0	2	0	1	1	3	0	8	Acetolactate synthase and RNA-directed DNA polymerase (Reverse transcriptase)
			7	4	2	11	2	13	0	2	0	38	

Table 32 continued. Proteins and single-nucleotide polymorphisms encoded on lineage-specific AF sequences

Lineage-specific AF	Size of AF (bp) in CC*	No. SNPs per AF (no. in ORFs)†	No. synonymous and non-synonymous SNPs in coding regions‡									No. of synonymous and non-synonymous SNPs§	ORFs within AF sequence¶
			CC1	CC5	CC8	CC22	CC30	CC45	CC59	CC97	ST93		
F9	389 (CC30)	28 (4)	0	0	0	0	0	1	0	0	0	1	Acetolactate synthase and RNA-directed DNA polymerase (Reverse transcriptase)
			0	3	0	0	0	0	0	0	0	3	
F10	479 (CC30)	16 (10)	1	0	0	0	1	0	2	0	0	4	Addiction module toxin, Txe/YoeB family and addiction module antitoxin, Axe family
			0	1	0	1	3	0	1	0	0	6	
F12	82 (CC45)	18 (15)	0	0	0	0	1	2	0	8	0	11	DNA Photolyase
			0	0	0	0	1	0	0	3	0	4	
F31	102 (CC45)	9 (9)	0	0	0	2	3	0	2	0	0	7	rRNA methyltransferase
			1	0	0	0	0	0	1	0	0	2	
F13	137 (CC45)	1 (1)	0	0	0	0	0	0	0	0	0	0	Abortive infection bacteriophage resistance protein
			0	0	1	0	0	0	0	0	0	1	
F33	317 (CC45)	3 (2)	0	0	0	0	0	0	0	0	0	0	Conserved hypothetical protein
			1	0	0	0	0	0	0	0	1	2	

Table 32 continued. Proteins and single-nucleotide polymorphisms encoded on lineage-specific AF sequences

Lineage-specific AF	Size of AF (bp) in CC*	No. SNPs per AF (no. in ORFs)†	No. synonymous and non-synonymous SNPs in coding regions‡									No. of synonymous and non-synonymous SNPs§	ORFs within AF sequence¶
			CC1	CC5	CC8	CC22	CC30	CC45	CC59	CC97	ST93		
F34	347 (CC45)	22 (14)	2	3	1	0	3	0	1	0	3	13	Acriflavin resistance transport protein
			0	0	0	0	0	0	0	0	1	1	
F35	354 (CC45)	26 (22)	4	0	3	1	1	0	0	0	7	16	Hydrolase (Esterase/Lipase)
			0	1	0	1	1	0	0	0	3	6	
F38	573 (CC45)	16 (10)	1	0	0	0	3	0	2	0	2	8	Haem biosynthesis related protein and Allergen V5/Tpx-1 related protein
			0	0	0	2	0	0	0	0	0	2	
F40	212 (CC59)	15 (13)	0	0	1	2	2	0	1	0	1	7	Dipeptidyl aminopeptidase/acylaminoacyl -peptidase
			0	1	1	0	3	0	0	0	1	6	
F41	222 (CC59)	12 (12)	0	0	1	7	3	0	1	0	0	12	Two-component response regulator BceR
			0	0	0	0	0	0	0	0	0	0	
F42	304 (CC59)	3 (3)	0	0	0	0	0	0	2	0	0	2	Bacteriophage portal protein
			0	0	0	0	0	0	1	0	0	1	

Table 32 continued. Proteins and single-nucleotide polymorphisms encoded on lineage-specific AF sequences

Lineage-specific AF	Size of AF (bp) in CC*	No. SNPs per AF (no. in ORFs)†	No. synonymous and non-synonymous SNPs in coding regions‡									No. of synonymous and non-synonymous SNPs§	ORFs within AF sequence¶
			CC1	CC5	CC8	CC22	CC30	CC45	CC59	CC97	ST93		
F15	340 (CC59)	12 (12)	1	0	0	2	2	0	0	0	0	5	Allophanate hydrolase 2 subunit 2
			1	0	1	1	0	0	3	0	1	7	
F43	350 (CC59)	21 (14)	2	3	1	1	3	0	1	0	3	14	Hypothetical protein and RND multidrug efflux transporter, Acriflavinresistance protein
			0	0	0	0	0	0	0	0	0	0	
F44	389 (CC59)	6 (6)	1	1	0	1	0	0	2	0	1	6	Murein hydrolase regulator LrgA and antiholin-like protein LrgB
			0	0	0	0	0	0	0	0	0	0	
F45	429 (CC59)	17 (15)	1	3	1	2	2	0	0	0	2	11	Coproporphyrinogen III oxidase
			0	0	2	1	0	0	0	0	1	4	
F46	447 (CC59)	48 (18)	0	4	0	1	0	0	1	0	0	6	Acetolactate synthase
			9	1	1	0	0	0	0	0	1	12	
F47	537 (CC59)	43 (43)	0	3	2	30	1	0	2	0	0	38	Nitrite reductase
			0	0	0	4	1	0	0	0	0	5	

Table 32 continued. Proteins and single-nucleotide polymorphisms encoded on lineage-specific AF sequences

Lineage-specific AF	Size of AF (bp) in CC*	No. SNPs per AF (no. in ORFs)†	No. synonymous and non-synonymous SNPs in coding regions‡									No. of synonymous and non-synonymous SNPs§	ORFs within AF sequence¶
			CC1	CC5	CC8	CC22	CC30	CC45	CC59	CC97	ST93		
F48	551 (CC59)	111 (95)	6	28	0	3	2	0	1	0	1	41	Lipoprotein
			4	41	0	1	3	0	1	0	4	54	
F49	571 (CC59)	19 (10)	1	0	0	0	3	0	2	0	2	8	Haem biosynthesis related protein and Allergen V5/Tpx-1 related protein
			0	0	0	2	0	0	0	0	0	2	
F50	599 (CC59)	12 (6)	0	2	0	1	0	0	1	0	2	6	Chromosomal replication initiator DnaA and DNA polymerase III, beta chain
			0	0	0	0	0	0	0	0	0	0	

*Size of amplified fragments (AFs) generated using FAFLP analysis with primers *Bgl*II+A and *Csp*6I+0 specific to one clonal complex (CC; shown in brackets).

†Number of single-nucleotide polymorphisms (SNPs), point mutations and indels in each AF. Numbers in brackets indicate the number of variations in ORFs of each AF.

[‡]Number of synonymous (top row) and non-synonymous (bottom row) variations, specific to one of nine CCs, in open reading frames (ORFs) of each AF.

[§]Number of synonymous and non-synonymous variations in ORFs of each AF.

[||]Proteins encoded on ORFs of each AF identified via BLAST searches against the NCBI database (http://www.ncbi.nlm.nih.gov/).

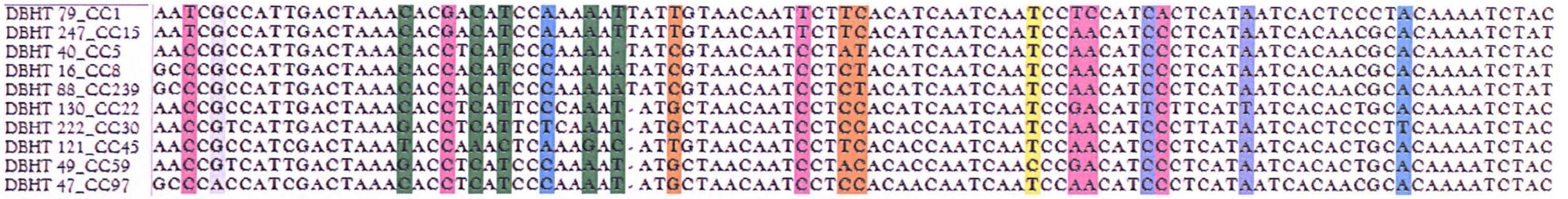
3.3.2.11. Pyrosequencing™

The 307 bp (F8) fragment specific to CC30 isolates was identified as a Pyrosequencing™ target with variations for eight lineages (Table 30). A 97 bp region within this AF was selected as a target for pyrosequencing. This region exhibited SNPs and point mutations specific to isolates assigned to one of seven lineages (CC1, 5, 22, 30, 45, 59 and 97). Pyrosequencing™ was performed on 23 isolates which were assigned to CC1, 5, 8, 15, 22, 30, 45, 59 and 672. Pyrosequencing™ amplified an amplicon in 18 of the 23 isolates (78 %). Alignment of these 18 sequences to the expected sequence revealed three sequences were less than 28 bp and could not be assigned to a lineage based on SNPs and point mutations (Appendix IX). The remaining 15 isolates were assigned to a lineage based on lineage-specific variations. The lineage assignment of these isolates correlated with MLST data (Figure 37).

3.3.2.12. Correlation of FAFLP data with *spa* typing and whole-genome mapping (WGM)

FAFLP data were compared amongst 18 and 74 isolates analysed using WGM and *spa* typing respectively. FAFLP profiles sampled between 0.53 % to 0.55 % of the genome amongst these isolates. WGM encompassed the entire genome and *spa* typing sampled approximately 0.02 % of the genome. Seventeen FAFLP profiles and 12 distinct optical maps were exhibited amongst the subset of 18 isolates and 69 FAFLP profiles were exhibited amongst the subset of 74 isolates which belonged to 39 *spa* types. The D and CI values based on FAFLP, *spa*, and WGM data were calculated for both subsets of isolates (Table 33).

Figure 37. Pyrosequencing™ region depicting single-nucleotide polymorphisms and point mutations



Target region (97 bp) of Pyrosequencing™ assay was selected based on 307 bp region of a CC30-specific amplified fragment generated using FAFLP analysis with primers *Bgl*I+A and *Csp*6I+0. Highlighted nucleotide positions display SNPs and point mutations specific to one of eight clonal complexes (CC). The combination of sequence variations highlighted in each colour were utilised to assign isolates to one of eight CCs. Positions in pink harbour CC1- or CC15-specific variations, grey harbour CC97, light green harbour CC45, light blue harbour CC30, orange harbour CC5, yellow harbour CC59 and light purple is CC22.

Table 33. Discriminatory power and confidence intervals amongst isolates analysed using FAFLP, *spa* typing and whole-genome mapping (WGM)

Analysis method	D value	CI value
FAFLP of isolate set A*	0.993	0.977 to 1.000
WGM of isolate set A*	0.922	0.833 to 1.000
FAFLP of isolate set B†	0.998	0.995 to 1.000
spa typing of isolate set B†	0.975	0.965 to 0.986

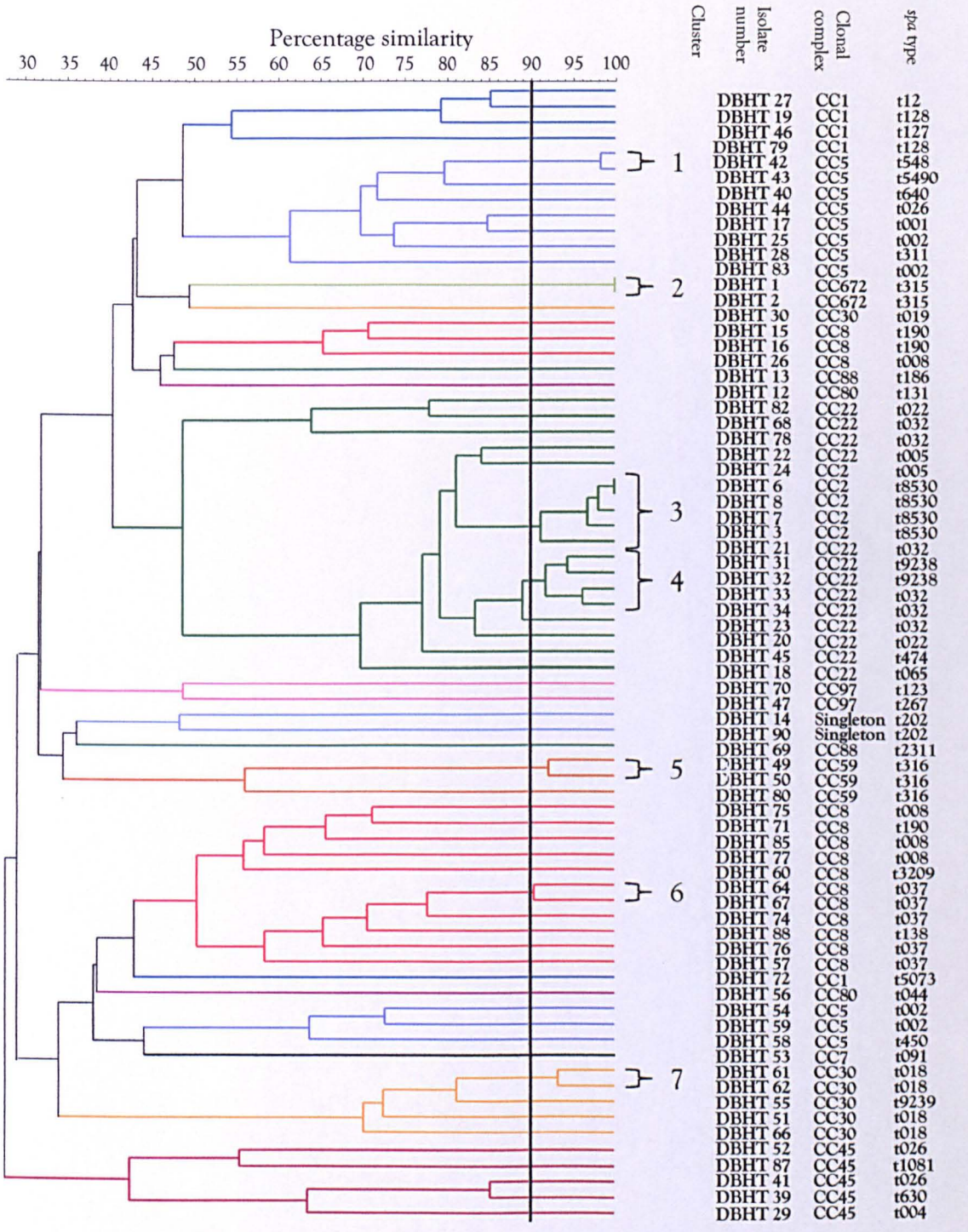
The discriminatory power (D) (Simpson, 1949) and confidence interval (CI) (Grundmann *et al.*, 2001) values were calculated based on FAFLP, *spa* typing and whole-genome mapping (WGM) data amongst two sets of isolates.

*Calculations were based on 18 isolates (set A) analysed using FAFLP and WGM.

†Calculations were based on 74 isolates (set B) analysed using FAFLP and *spa* typing.

Comparison of dendrograms based on FAFLP and *spa* typing data revealed isolates in each cluster were assigned to the same lineage. FAFLP identified greater heterogeneity (percentage difference) between isolates assigned to the same lineage as compared to *spa* typing. Based on a 90 % similarity cut-off applied to the dendrograms, 5 % of isolates (n=4) displayed divergent *spa* profiles (Figure 26) whilst 74 % of isolates (n=70) exhibited divergent FAFLP profiles (Figure 38). Cluster analysis based on FAFLP and WGM data grouped isolates assigned to CC22, 30 and 45 in distinct clusters (see WGM: section 3.4.4). However, the heterogeneity identified amongst these isolates varied between the two analysis methods. FAFLP analysis identified 31.0 %, 57.0 % and 63.0 % divergence whilst WGM identified 6 %, 7.5 % and 13.0 % difference amongst CC22, 30 and 45 isolates respectively.

Figure 38. Cluster analysis using UPGMA on FAFLP data



The dendrogram was derived from FAFLP data for 73 isolates using UPGMA cluster analysis. Based on 90 % similarity cut-off (shown as vertical black line), seven FAFLP clusters were identified (shown as black numbers). The clonal complex (CC) and *spa* type of isolates is shown on the right of the cluster number. FAFLP identified the heterogeneity between isolates of each lineage.

The number of lineage-specific fragments identified using FAFLP and WGM were compared (Table 34). Lineage-specific fragments identified using WGM covered a larger percentage of the genome than FAFLP. Fragments specific to CC1, 5 and 8 were identified using FAFLP whilst WGM identified fragments specific to CC239. In both cases the second method (FAFLP or WGM) failed to identify fragments specific to those respective lineages.

3.3.3. Insertion sequence targeted fluorescent amplified fragment length polymorphism (IS-FAFLP)

Searches for an IS present within WGSs of *S. aureus* was performed. Initial searches based on NCBI database annotations of 15 WGSs did not identify any ISs from the same family present in all sequences. Comparisons of these 15 sequences using Mauve sequence alignment software revealed no sequences that met the search terms 'insertion' and 'sequence' amongst 11 of the 15 WGSs. Two regions (131 bp and 783 bp sequences) were identified within the strain MRSA252 which encoded putative IS proteins but were probable pseudogenes whilst the *tnp* gene which encodes for the transposase (1319 bp) associated with IS*1181* was identified in three strains (Mu3, Mu50 and N315). Search results using the keyword 'transposase' yielded between seven and 46 sequences amongst each of the 15 genomes (Appendix X). BLAST search results of these transposase sequences revealed similar sequences were present in up to 12 WGSs. However, no transposase sequence displayed similar sequences in each of the 15 genomes. The transposase sequence (T1) from the strain ED98 (981105 to 982280 bp) was selected as a FAFLP target as similar sequences were identified in 12 of the 15 genomes. The consensus sequence (1320 bp) of this region was conserved (99.16 %). However, no similar sequences were identified amongst three strains (MSSA476, MW2 and RF122).

Table 34. Lineage-specific fragments identified using FAFLP and WGM

	FAFLP			WGM		
CC*	No. lineage-specific fragments†	Size of genome represented (bp)‡	Size of genome represented (%)‡	No. lineage-specific fragments†	Size of genome represented (bp)‡	Size of genome represented (%)‡
CC1	1	355	0.01 %	N/A§	N/A	N/A
CC5	4	1456	0.05 %	N/A	N/A	N/A
CC8	2	642	0.02 %	N/A	N/A	N/A
CC239	N/A	N/A	N/A	3 (1)	30000	1.05 %
CC22	6	1377	0.05 %	14 (3)	212500	7.46 %
CC30	13	3628	0.13 %	7 (2)	140000	4.91 %
CC45	11	3655	0.13 %	5 (1)	119000	4.18 %
CC59	13	5036	0.18 %	15 (1)	245000	8.60 %

*CC, clonal complex.

†Lineage-specific fragments identified using FAFLP and whole-genome mapping (WGM). Brackets indicate number of lineage-specific regions.

‡The size of the genome represented by the lineage-specific fragments identified using FAFLP or WGM are shown in base pairs and as a percentage of the genome.

§N/A, not applicable as fragments specific to the lineage were not identified using FAFLP or WGM.

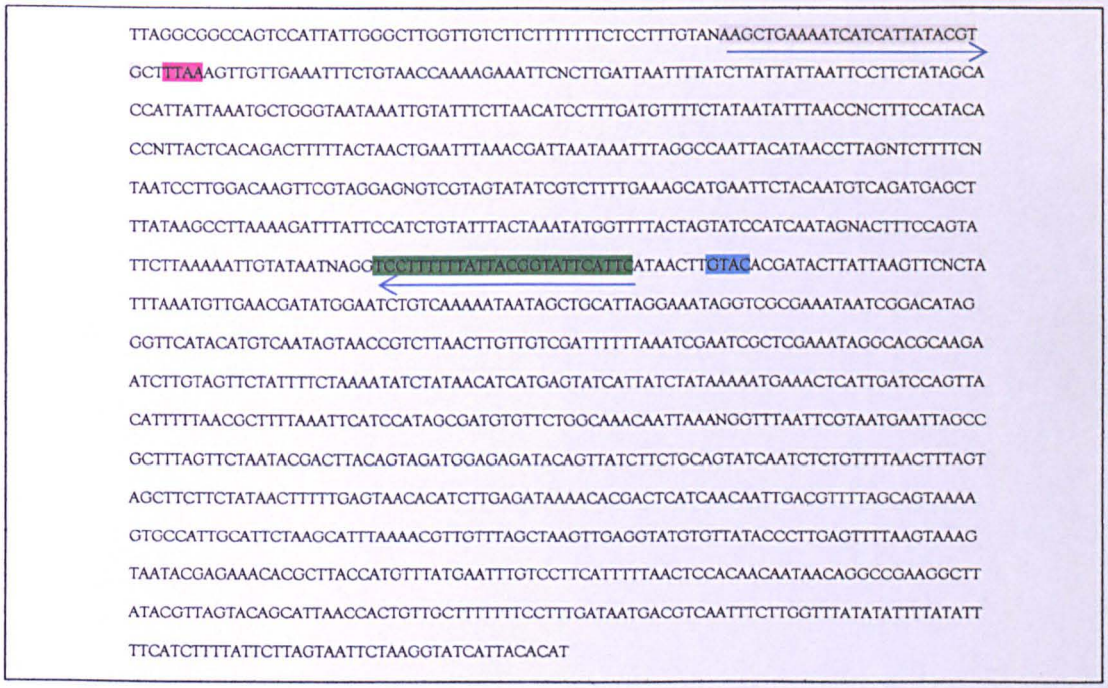
A second search was performed for an additional FAFLP target which may enable the application of this assay to the majority of *S. aureus* isolates. The transposase sequence (T2) from the strain MRSA252 (2306521 to 2307052 bp) was selected as similar sequences were identified amongst eight strains which included MSSA476 and MW2. This consensus sequence (534 bp) was also relatively conserved (84.63 %). Sequences similar to T1 and/or T2 were identified in 14 WGSs. However, sequences similar to T1 and T2 were not identified within the strain RF122.

Restriction endonuclease recognition sequences within the transposase sequences (T1 and T2) were analysed using GeneQuest™. This identified a recognition sequence for the endonucleases *Mse*I and *Csp*6I within the T1 sequence and a *Csp*6I recognition sequence within the T2 sequence. Reverse primers were designed manually upstream of the restriction endonuclease recognition sequences (Figure 39) (Figure 40). The primer sequences were 5' CACGTATAATGATGATTTTCAGCTT 3' for T1_*Mse*I, 5' GAATGAATACCGTAATAAAAAAGGA 3' for T1_*Csp*6I and 5' TCTTGGCTTATACATGACTATTTTC 3' for T2_*Csp*6I. The primer name represents the transposase sequence (T1 or T2) targeted by the reverse primer and the restriction endonuclease utilised in FAFLP. An IS-FAFLP assay targeting T1 was initially performed on DBHT 1 to 24 from SRL using the endonuclease *Mse*I for DNA digestion. Fragments were amplified using the *Mse*I+0 (Table 9) and T1_*Mse*I primers and resolved on a 2 % agarose gel. Between one to three fragments less than 200 bp in size were amplified amongst these isolates.

Further IS-FAFLP assays targeting T1 and T2 were performed on isolates DBHT 1 to 24 from SRL using the endonuclease *Csp*6I for DNA digestion. Fragments were amplified using

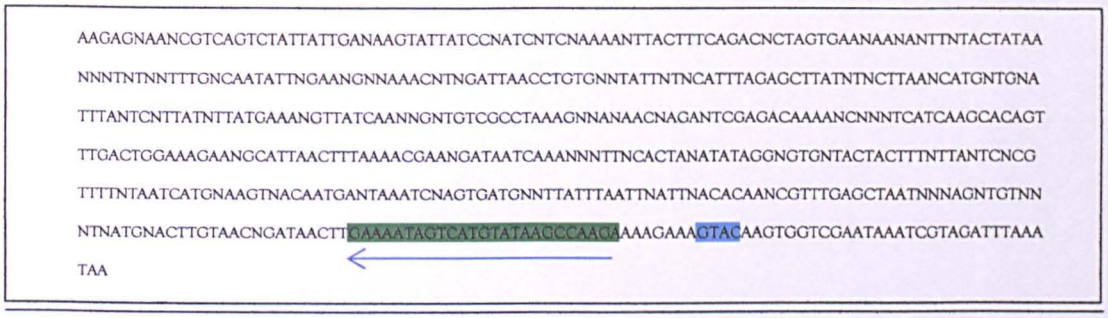
the *Csp*6I+0 (Table 9) and either the T1_*Csp*6I or T2_*Csp*6I primer and resolved on a 2 % agarose gel. All isolates excluding one (DBHT 7 which yielded 16 fragments) amplified a large number of fragments which could not be accurately differentiated.

Figure 39. Insertion-sequence FAFLP target transposase sequence 1



Consensus sequence of the transposase 1 (T1) target utilised in an insertion-sequence targeted FAFLP assay. Highlighted nucleotides in purple and blue indicate *Mse*I and *Csp*6I restriction endonuclease recognition sequences respectively. Nucleotides highlighted in grey and green indicate the T1_*Mse*I and T1_*Csp*6I reverse primer annealing sequences respectively. Arrows indicate the orientation of the reverse primer.

Figure 40. Insertion-sequence FAFLP target transposase sequence 2



Consensus sequence of the transposase 2 (T2) target utilised in an insertion-sequence targeted FAFLP assay. Highlighted nucleotides in blue indicate a *Csp*6I restriction endonuclease recognition sequence. Nucleotides highlighted in green indicate the T2_*Csp*6I reverse primer annealing sequence. Arrows indicate the orientation of the reverse primer.

IS-FAFLP PCRs were performed on DBHT 9 to 16 from SRL using the transposase-specific reverse primers (T1_*Csp*6I and T2_*Csp*6I) on *Eco*RI with *Mse*I- and *Hin*dIII with *Hha*I-digested DNA. As the reverse primers T1_*Csp*6I and T2_*Csp*6I were transposase-specific the FAFLP AFs should harbour partial T1 or T2 sequences. In addition, the use of a double digest reaction generated less fragments as compared to when a single frequent-cutting endonuclease was utilised. Isolates analysed with the majority of IS-FAFLP assays yielded less than five fragments when resolved on a 2 % agarose gel. However, AFs generated using the primer T2_*Csp*6I on *Hin*dIII with *Hha*I digested DNA yielded a range of fragments between 50 to 600 bp which could be discriminated on an agarose gel. This IS-FAFLP PCR was repeated for DBHT 1 to 15. Isolates analysed using the T1_*Csp*6I primer amplified no fragments. Each of the 15 profiles generated using the primer T2_*Csp*6I displayed background fluorescence of greater than 100 RFU, therefore the number of

fragments were not distinguishable. Therefore no IS-FAFLP assay produced FAFLP profiles which could distinguish between isolates.

3.3.4. WGM

Of the 18 isolates analysed using WGM, six were MSSA and were representative of CC8, 15, 22 and 30. The remaining 12 isolates were MRSA and were assigned to CC1, 5, 8, 22, 30, 45 and 59. A single isolate represented the majority of CCs with the exception of CC8 (n=3) and CC45 (n=2). MRSA isolates belonged to SCC*mec* types I to VI which included four SCC*mec* type IV subtypes and one non-subtypeable element. Each optical map was composed of 192 to 227 fragments, each of which was between 1174-94098 bp in size (Table 35).

Cluster analysis was performed using the UPGMA method on WGM data for 18 isolates and *in silico* data for 21 strains obtained from the NCBI database. Optical maps were concordant with MLST data. Based on MLST CCs, a 90 % similarity cut-off was applied to the dendrogram (Figure 41). The 39 entries were grouped into six clusters and six maps displayed greater than 10 % divergence. Cluster 1 isolates (CC22) displayed up to 18.4 % difference compared to isolates from other clusters. Isolates belonging to CC30 (cluster 4) displayed up to 23.9 % divergence from other isolates. Cluster 6 isolates (CC59) displayed the greatest diversity (up to 36.9 %) with isolates of other lineages.

Cluster 1 was composed of seven CC22 entries, one of which was an *in silico* optical map based on a preliminarily annotated EMRSA-15 strain. WGM identified heterogeneity within the CC22 isolates. DBHT 18 (CC22-II) displayed the greatest similarity with EMRSA-15. The remaining isolates displayed up to 5.3 % divergence from this strain and other CC22

maps. DBHT 4 and 5 were MSSA and displayed indistinguishable optical maps. The remaining four CC22 MRSA isolates displayed different *spa* types. WGM identified heterogeneity between all optical maps.

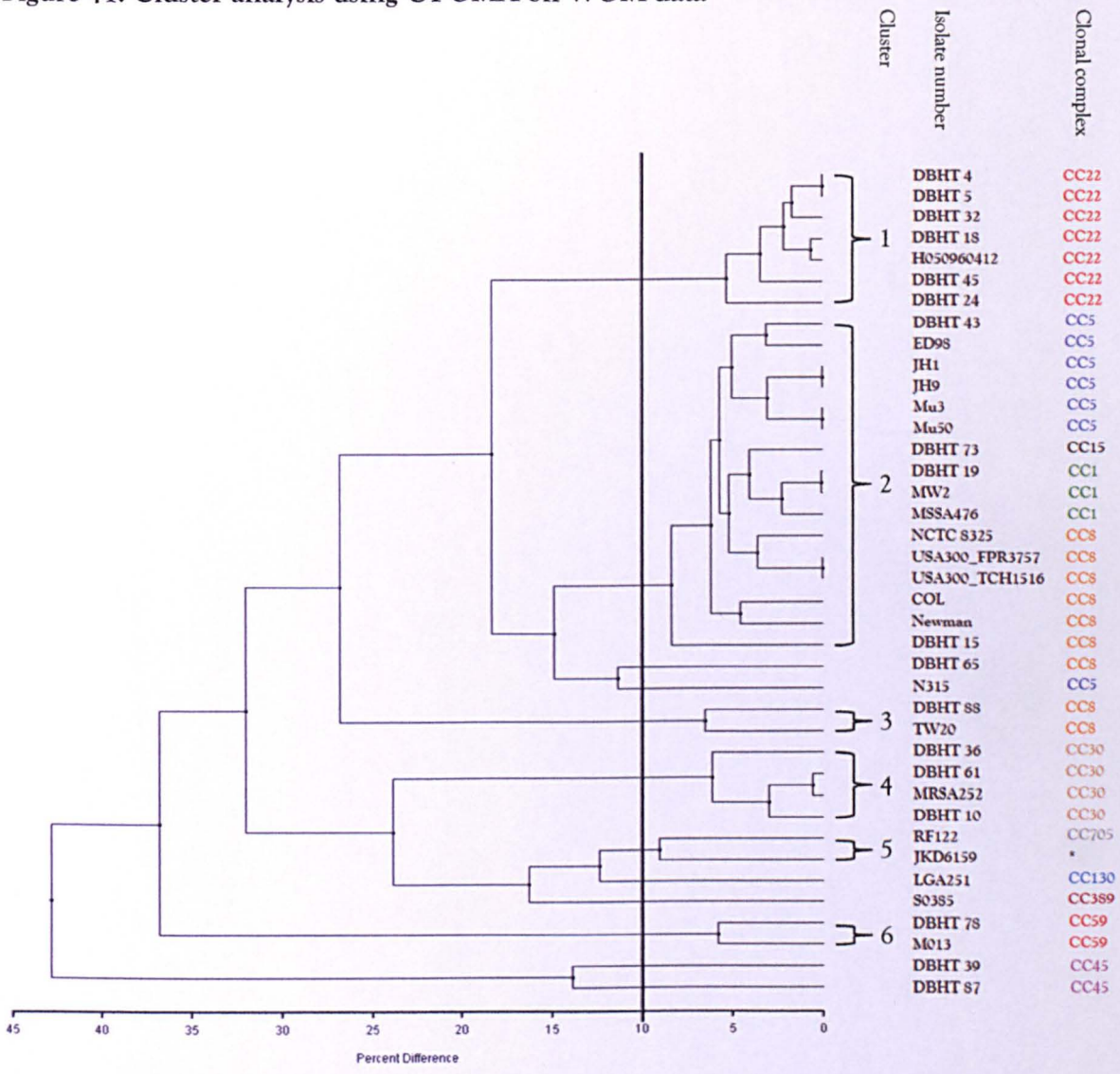
Table 35. Characteristics of optical maps generated using WGM

Isolate number	Minimum fragment size (bp)*	Maximum fragment size (bp)*	No. of fragments†
DBHT 4	1756	70027	212
DBHT 5	1625	70417	216
DBHT 10	1487	63863	215
DBHT 15	1174	90716	227
DBHT 18	1275	69646	217
DBHT 19	1595	94098	219
DBHT 24	1818	81560	205
DBHT 32	1480	70631	218
DBHT 36	1549	63137	201
DBHT 39	1237	67788	201
DBHT 43	1487	89174	212
DBHT 45	1576	68682	205
DBHT 61	1536	53268	227
DBHT 65	1454	92276	211
DBHT 73	1500	90852	205
DBHT 87	1434	62595	220
DBHT 78	1638	81474	192
DBHT 88	1657	83210	215

*Display minimum and maximum size of fragments amongst 18 optical maps generated using whole-genome mapping (WGM) with the endonuclease *NocI*.

†Number of fragments constituting each optical map.

Figure 41. Cluster analysis using UPGMA on WGM data



The dendrogram was derived from whole-genome mapping (WGM) data using the restriction endonuclease *Noc*I for 18 isolates. Based on a 90 % similarity cut-off (shown as vertical black line), six WGM clusters were identified (shown as black numbers). The clonal complex (CC) assignment of isolates and isolate number are shown on the right. Isolates assigned to each CC are highlighted in a single colour.

***Isolate assigned to a singleton ST.**

Cluster 2 consisted of 16 entries assigned to CC1, 5, 8 and 15. Based on a further 95 % similarity cut-off the isolates composed four sub-clusters and a single isolate. One sub-cluster

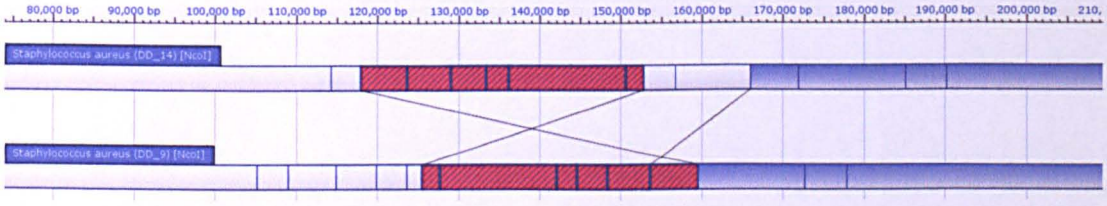
consisted of six CC5 entries, five of which were *in silico* optical maps. DBHT 43 was most closely related (3.1 % difference) to the strain ED98 and exhibited up to 5.1 % divergence from the remaining CC5 WGSs. A second sub-cluster was composed of a CC15 and three CC1 entries two of which were generated from WGSs. DBHT 73 (CC15) displayed up to 4.1 % divergence from the remaining CC1 entries. One CC1 isolate (DBHT 19) displayed an optical map indistinguishable to the strain MW2. The remaining two sub-clusters were composed of WGSs assigned to CC8. Indistinguishable maps were generated for two USA300 strains which exhibited 3.6 % divergence from the strain NCTC 8325. Optical maps based on strains COL and Newman revealed 4.6 % difference. Isolate DBHT 15 (CC8) was divergent from the remaining isolates in the cluster and exhibited up to 8.4 % difference from the other entries.

The two ST239 (CC8) isolates constituting cluster 3 exhibited 6.6 % difference from each other. Cluster 4 was composed of four CC30 isolates and one MRSA252 (EMRSA-16) strain. DBHT 61 exhibited 99.4 % similarity with MRSA252. The other two MSSA CC30 isolates (DBHT 10 & 36) exhibited up to 6.2 % divergence from the other entries. Cluster 5 consisted of two whole-genome sequenced strains which exhibited up to 9.1 % difference. One of these strains was a bovine strain whilst the other was an Australian CA-MRSA clone. Two CC59 entries formed cluster 6 and DBHT 78 exhibited up to 5.8 % difference as compared to the whole-genome sequenced strain M013.

Of the six optical maps which displayed greater than 10 % divergence from the remaining isolates, two were assigned to CC45 and two belonged to clones associated with live-stock. The two remaining isolates, DBHT 65 and the strain N315, belonged to CC8 and 5 respectively but displayed divergence from other isolates assigned to the same CC.

Multiple alignment on the 39 optical maps revealed insertions, deletions and an inversion between genomes. Optical maps displayed greater similarity between isolates assigned to the same CC. Fragment patterns unique to isolates from a single CC were identified and one inversion was identified among the maps. The inversion consisted of a region (34000 to 37500 bp) displaying six or seven fragments which were present in all isolates. However, the optical map for isolate DBHT 24 (CC22) displayed this fragment pattern in a reverse orientation. Four fragments immediately to the left of this inverted region in DBHT 24 formed a similar pattern to those of the first four fragments of this inverted region in all other isolates (Figure 42). Comparison to *in silico* WGSs revealed this inverted region encoded numerous membrane and cell surface proteins including the *spa* gene.

Figure 42. Optical maps displaying inversion between CC22 isolates

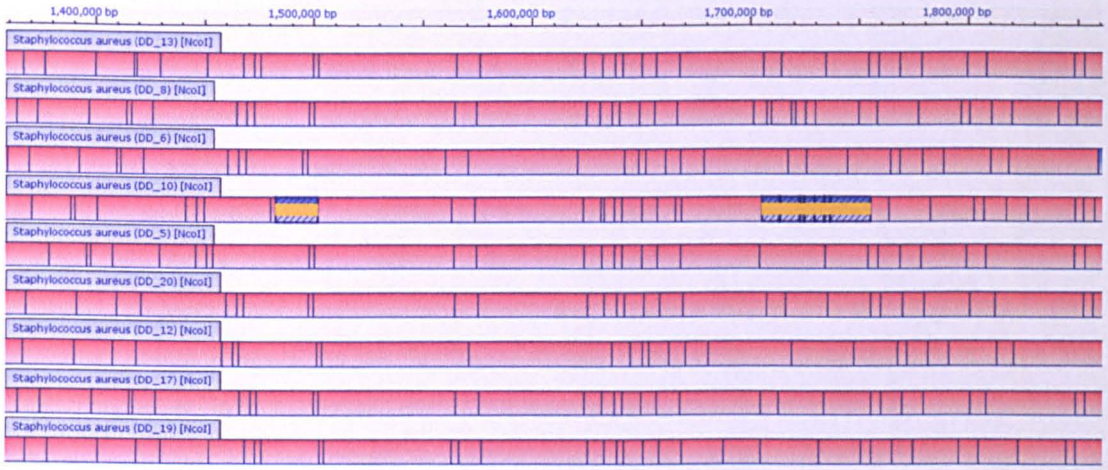


DD_14, DBHT 45 (CC22); DD_9, DBHT 24 (CC22). The figure depicts optical maps with *Nod*I restriction endonuclease recognition sites (vertical bars). Highlighted region in red displays a restriction fragment pattern in the reverse orientation in DBHT 24 as compared to other isolates such as DBHT 45 indicating an inversion has occurred in this region in isolate DBHT 24. Highlighted region in blue indicate regions similar between the optical maps whilst those in grey are unique to each map.

Restriction fragment patterns unique to isolates assigned to a single CC were identified. Patterns unique to isolates assigned to CC22, 30, 45, 59 and ST239 were also identified.

Lineage-specific restriction fragment patterns exclusive to isolates assigned to CC1, 5 or 8 were not found. Of the six CC22 isolates, four isolates (DBHT 4, 5, 18 and 32) displayed three unique regions. The first CC22-specific region was identified 1360 kb into the sequence and was composed of a single fragment approximately 19 kb in size, comparison of this region to the *in silico* map of the strain H050960412 revealed this region contained genes encoding proteins involved in the repair of damaged DNA, an ion channel and an IS protein (Figure 43).

Figure 43. Multiple alignment of optical maps displaying CC22-specific fragment patterns



DD_13, DBHT 43 (CC5); DD_8, DBHT 19 (CC1); DD_6, DBHT 15 (CC8); DD_10, DBHT 32 (CC22); DD_5, DBHT 10 (CC30); DD_20, DBHT 88 (CC8); DD_12, DBHT 39 (CC45); DD_17, DBHT 73 (CC15); DD_19, DBHT 78 (CC59). The figure depicts optical maps with *NocI* restriction endonuclease recognition sites (vertical bars). Highlighted regions in red indicate regions of similarity between multiple optical maps. Highlighted regions in blue with yellow bars indicate regions specific to a subset of CC22 isolates.

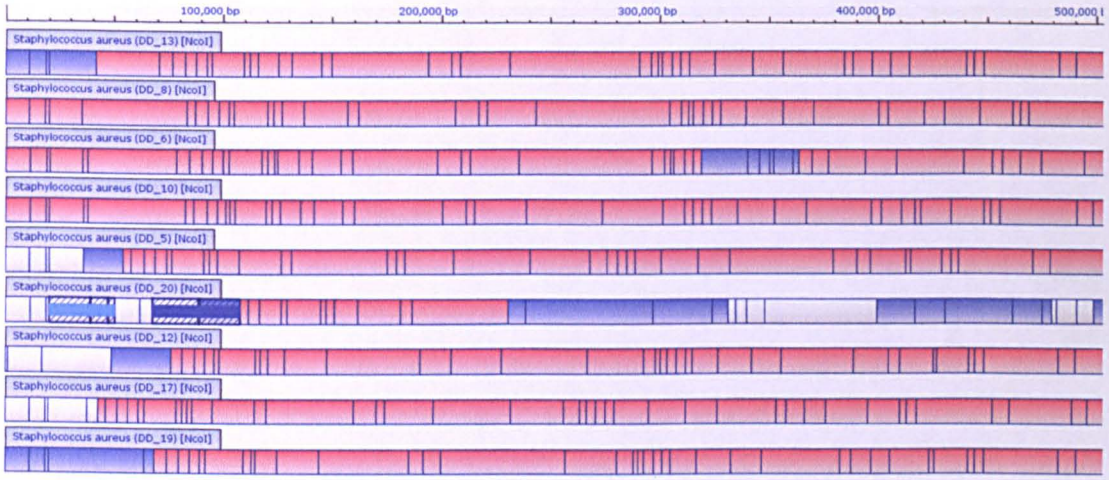
A second unique region was located 1600 kb into the sequence and was formed of seven fragments ranging in size from 2 to 18 kb. Features within this region included bacteriophage proteins, a transposase and integrase/recombinase (Figure 43). The final fragment pattern consisted of six fragments found at the end of the map which covered 144 kb. This region harboured genes encoding for surface proteins, a transposase and conserved proteins.

A single region unique to ST239 approximately 30 kb in size was identified and was formed of three fragments 20 kb into the genome. Comparison with the *in silico* optical map of the strain TW20 revealed that this region encodes for proteins associated with the SCC*mec* such as *ccr*C, IS*431* and the mercury resistance operon (Figure 44). Two fragment patterns unique to isolates of CC30, in addition to ST239, were also identified. The first of these regions consisted of two fragments 67 kb into the genome and were approximately 40 kb in size. Comparison of this region with *in silico* optical maps from strains MRSA252 and TW20 revealed this region encoded a cadmium efflux system, restriction endonuclease and a partial SCC*mec* region which included the transposon Tn*554*, *ccr*A and *ccr*B genes (Figure 44). Five fragments at the end of the optical maps were also unique to CC30 and ST239 isolates. This region covered over 100 kb and encoded numerous surface proteins, membrane transporters and a transposase.

Lineage-specific restriction fragment patterns unique to CC45 and 59 were identified at the terminal end of optical maps. These regions covered up to 245 kb and were formed of 5 to 15 fragments (Figure 45). The CC45 optical map could not be compared with a WGS representative of CC45. The CC59-specific region was compared against the strain M013 and encoded features similar to those found at the terminal end of CC22 and 30 isolates. In

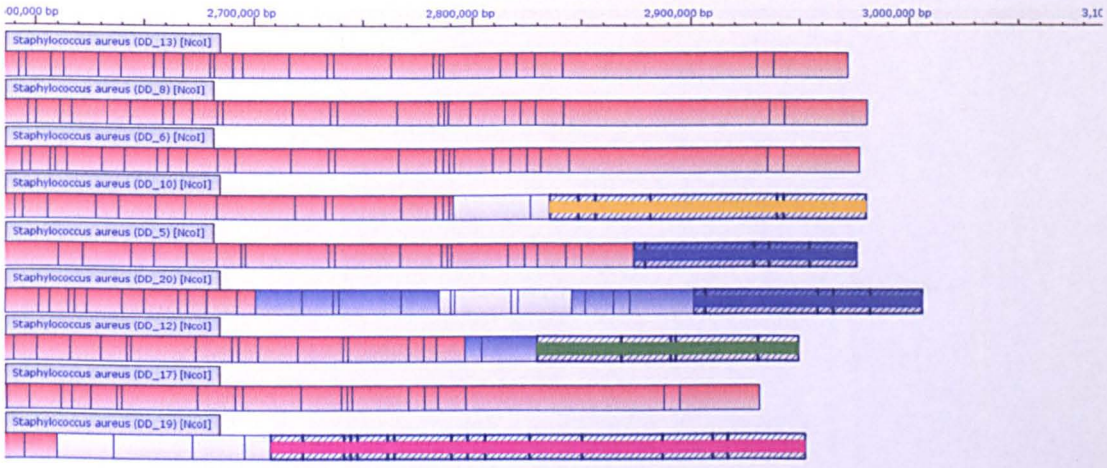
addition, numerous transposases with homology to IS*1272* and IS*605*/IS*200*, holin-like protein *cidA*, exotoxin G, antimicrobial transport proteins, numerous surface-associated proteins and a biofilm operon were encoded in this region.

Figure 44. Multiple alignment of optical maps showing lineage-specific regions at the beginning of maps



DD_13, DBHT 43 (CC5); DD_8, DBHT 19 (CC1); DD_6, DBHT 15 (CC8); DD_10, DBHT 32 (CC22); DD_5, DBHT 10 (CC30); DD_20, DBHT 88 (CC8); DD_12, DBHT 39 (CC45); DD_17, DBHT 73 (CC15); DD_19, DBHT 78 (CC59). The figure depicts optical maps with *NocI* restriction endonuclease recognition sites (vertical bars). Highlighted regions in red indicate regions of similarity between multiple optical maps. Highlighted regions in blue indicate regions of similarity between two optical maps. Highlighted regions in grey indicate regions unique to a single optical map. Highlighted regions in light blue bars indicate regions specific to ST239 and dark blue bars indicate regions specific to CC30 and ST239 isolates.

Figure 45. Multiple alignment of optical maps showing lineage-specific regions at terminal end of maps



DD_13, DBHT 43 (CC5); DD_8, DBHT 19 (CC1); DD_6, DBHT 15 (CC8); DD_10, DBHT 32 (CC22); DD_5, DBHT 10 (CC30); DD_20, DBHT 88 (CC8); DD_12, DBHT 39 (CC45); DD_17, DBHT 73 (CC15); DD_19, DBHT 78 (CC59). The figure depicts optical maps with *NocI* restriction endonuclease recognition sites (vertical bars). Highlighted regions in red indicate regions of similarity between multiple optical maps. Highlighted regions in blue indicate regions of similarity between two optical maps. Highlighted regions in grey indicate regions unique to a single optical map. Highlighted regions in dark blue bars indicate regions specific to CC30 and ST239 isolates. Highlighted regions in yellow bars indicate regions specific to CC22 isolates. Highlighted regions in green bars indicate regions specific to CC45 isolates. Highlighted regions in purple bars indicate regions specific to CC59 isolates.

WGM also identified the heterogeneity within isolates of a single CC. Multiple alignment of optical maps revealed fragment patterns unique to a single isolate. These restriction patterns could not be compared to *in silico* optical maps, therefore the features encoded in these regions were difficult to ascertain. However, these fragments and possible features encoded in these regions were noted (Table 36). The majority of unique regions were identified in CC8 isolates. In addition, the majority of isolate-specific regions were identified at the start and end of the optical maps.

A multiple alignment between an ST8, ST239 and ST30 isolate revealed 240 kb near the beginning of the ST239 optical map and the final 160 kb (13 %) displayed restriction fragment patterns similar to those found in ST30. The central 2280 kb region (76 %) displays restriction fragment patterns similar to ST8 isolates of which 1690 kb was also similar to ST30. The remainder of the ST239 optical map (up to 240 kb) was unique and did not display similarity to ST8 or ST30 (Figure 46).

Table 36. Isolate-specific fragments identified using WGM

Isolate number	CC*	No. of isolate-specific fragments†	Loci of isolate-specific fragments (bp)‡	Compared to strain‡	Notes§
DBHT 45	CC22	4	1-48918	H050960412	First two restriction-sites in *purA* and *yycJ*, second two differ from other isolates
DBHT 73	CC15	4	1-37672	N/A¶	No CC15 WGS available for comparison
DBHT 15	CC8	1	1777763-1791951	NCTC 8325, USA300, USA300_TCH1516, COL, Newman, TW20	Fragment not found in any *in silico* optical map
DBHT 15	CC8	1	2051880-2094702	NCTC 8325, USA300, USA300_TCH1516, COL, Newman, TW20	Fragment may encode bacteriophage phiNM1 proteins including PVL-like protein. Restriction-site in bacteriophage integrase sequence
DBHT 65	CC8	2	1-39049	NCTC 8325, USA300, USA300_TCH1516, COL, Newman, TW20	Two restriction-sites in *purA* and *yycJ*
DBHT 65	CC8	1	170746-191964	NCTC 8325, USA300, USA300_TCH1516, COL, Newman, TW20	Fragments not found in any *in silico* optical map

Table 36 continued. Isolate-specific fragments identified using WGM

Isolate number	CC*	No. of isolate-specific fragments†	Loci of isolate-specific fragments (bp)‡	Compared to strain‡	Notes§
DBHT 65	CC8	3	2697834-2830865	NCTC 8325, USA300, USA300_TCH1516, COL, Newman, TW20	Fragments not found in any *in silico* optical map
DBHT 36	CC30	3	182091-211211	MRSA252	Fragments not found in any *in silico* optical map
DBHT 61	CC30	8	1-17161	MRSA252	Two restriction-sites in *yycJ* and *sasH* replaced with one, two cut-sites in *ccrB* replaced with one
DBHT 61	CC30	1	532369-552622	MRSA252	Fragment may encoderibosomal proteins
DBHT 88	CC8	9	1-88704	TW20	Three fragments unlike *in silico* optical map this region included the SCC*mec*
DBHT 88	CC8	2	481585-498531	TW20	Region includes membrane proteins and two less restriction-sites compared to *in silico* optical map
DBHT 78	CC59	4	2468284-2566381	M013	Fragments not found in any *in silico* optical map
DBHT 39	CC45	3	1-48649	N/A	No CC45 WGS available for comparison
DBHT 87	CC45	5	1-62262	N/A	No CC45 WGS available for comparison
DBHT 87	CC45	1	2241108-2242542	N/A	No CC45 WGS available for comparison
DBHT 87	CC45	6	2458414-2564531	N/A	No CC45 WGS available for comparison

Table 36 continued. Isolate-specific fragments identified using WGM

Isolate number	CC*	No. of isolate-specific fragments†	Loci of isolate-specific fragments (bp)‡	Compared to strain‡	Notes§
DBHT 87	CC45	7	2761176-2844901	N/A	No CC45 WGS available for comparison

*CC, clonal complex.

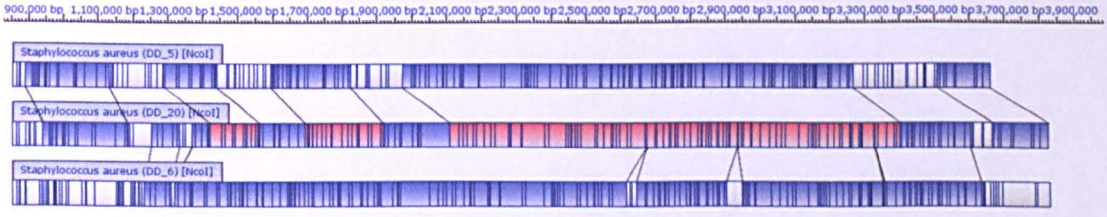
†Number of fragments specific to a particular isolate.

‡Loci of isolate-specific fragments relative to the whole-genome sequence (WGS) of the strain displayed.

§Features encoded in these regions relative to the WGS of the strain displayed.

||N/A, not applicable.

Figure 46. Multiple alignment of optical maps from ST8, ST30 and ST239 isolates



DD_5, DBHT 10 (CC30); DD_20, DBHT 88 (CC8); DD_6, DBHT 15 (CC8). Highlighted regions in red indicate regions of similarity between multiple optical maps. Highlighted regions in blue indicate regions of similarity between two optical maps. Highlighted regions in grey indicate regions unique to a single optical map.

CHAPTER 4

4. Discussion

Molecular methods are useful in microbial diagnostics, epidemiology and research. In particular, well-established molecular methods such as MLST and SCC*mec* typing have been invaluable in understanding the population structure and evolution of *S. aureus*, a human pathogen. The bacterium is a common cause of HA infection in immunocompromised patients, individuals with open wounds and the elderly. However, infections affecting the healthy and young have increased in the past few decades and are a rising clinical concern. Understanding the genetic diversity of *S. aureus* can therefore result in a better understanding of the pathogenesis, spread and evolution of *S. aureus*, which in turn can lead to the development of effective infection control measures. In the present study, molecular methods were utilised to investigate differences in the distribution and variability of bacterial genes present between successful and unsuccessful lineages of MRSA. Furthermore, this study explored these genetic differences in relation to functional differences within the bacterium, which could provide a selective advantage in successful lineages, for example, by altering their virulence or pathogenicity.

The BSAC collection provided a well-characterised set of isolates with data for MIC (for six antimicrobials), *mec*A and *mup*A PCR and the number of isolates collected from each centre available. As MIC testing for all isolates was performed by a single laboratory, this eliminated the inter-laboratory variability of these data. Since bacteraemia is a systemic infection, analysing bacteraemia isolates provided an advantage as compared to isolates from other sites of infection where it may be difficult to differentiate between those isolates causing infection and colonising isolates. Studying bacteraemia isolates could help us understand *S. aureus* infections at other sites. However, bacteraemia isolates may not be representative of the entire *S. aureus* population; they may represent a subset such as those resistant to treatment at the initial site of infection. The BSAC isolates were also selected as they were

representative of urban and rural settings, with a range of deprivation scores which can be used as a measure of the standard of living in an area. This may help reduce sample bias, however, it should be noted data on the geographical locations throughout England and Ireland from which the isolates were collected was not available. As isolates were collected consecutively as opposed to randomly throughout the year, and due to the infectious nature of bacteria, the independence of each isolate was reduced, although duplicate isolates from the same patient were not collected (Reynolds *et al.*, 2008; 2009). Statistically significant and meaningful results in this study were drawn from results obtained using this collection of isolates.

The SRL collection was representative of different centres however, their geographic locations were unavailable. Although the epidemiological data for this collection was limited, it included multiple isolates from the same centre including those obtained from a single patient. Preliminary FAFLP analysis of this collection determined FAFLP as a molecular tool to highlight small-scale sequence variations (point mutations and indels) within isolates from a single patient. Furthermore, this collection assisted the development and optimisation of a novel FAFLP method which was subsequently applied to all isolates in this study. WGS data were available for six strains which facilitated *in silico* analysis on strains and a direct comparison with experimental data for the same strains. The Glasgow collection represented a well-characterised isolate collection for which MIC, *nuc*, *mec*A and *mup*A PCR and PFGE information was available. In addition, inclusion of this collection from Scotland ensured that the study represented isolates from different regions of the UK. The CMPHL collection included isolates involved in an outbreak and isolates which colonised patients and staff during this period. FAFLP analysis of this collection enabled the author to assess the genetic

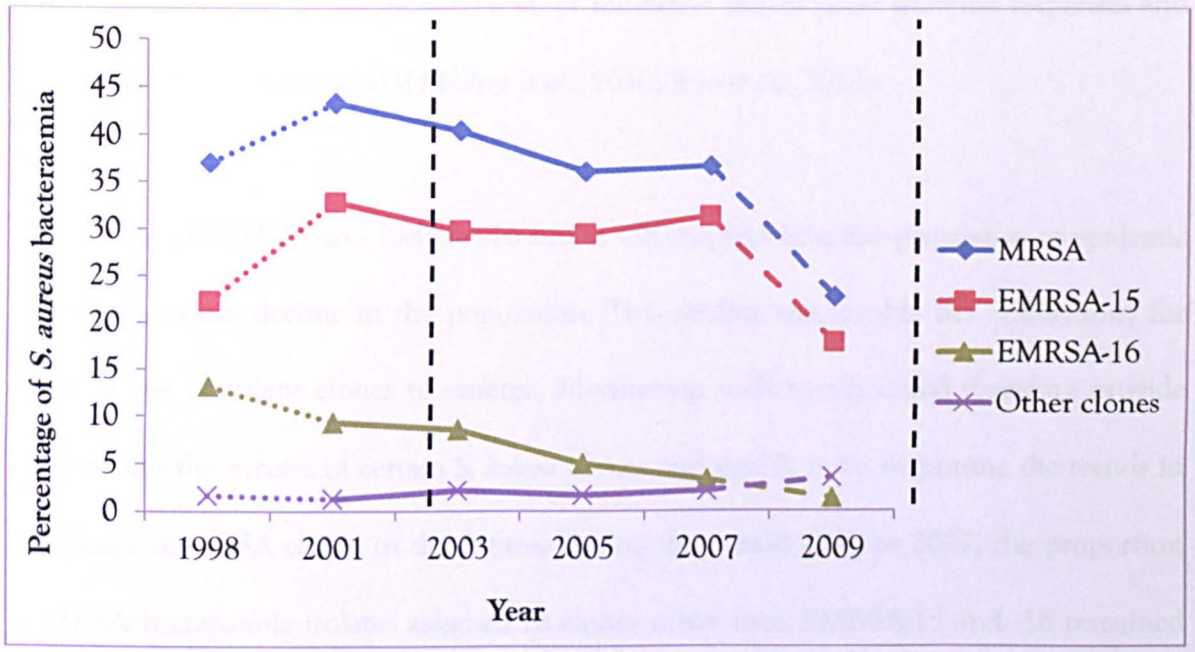
diversity between outbreak and colonising isolates and the ability of FAFLP to differentiate between these groups of isolates.

4.1. Epidemiology of study isolates

The proportion of cases of *S. aureus* bacteraemia caused by MRSA peaked and subsequently declined during the 2000s (43 % in 2001, 23 % in 2009 and 14 % in 2011) (Hope *et al.*, 2008). During the period 1998 to 2007, the majority of MRSA bacteraemia was attributable to two strains (ST22-MRSA-IV or EMRSA-15 and ST36-MRSA-II or EMRSA-16). This study has built on that previous molecular epidemiological surveillance of MRSA causing bacteraemia by using conventional typing tools and has developed new tools to probe the genetic diversity of MRSA in greater depth.

The current study identified that the majority of cases of MRSA bacteraemia in 2009 continued to be attributed to these two highly successful strains. Cases due to EMRSA-15 and EMRSA-16 continued to fall from 31 % and 3 % in 2007 to 18 % and 1 % of the *S. aureus* bacteraemia collection in 2009 respectively (Figure 47). This suggested that against a backdrop of a MRSA bacteraemia rate that has peaked and is now falling, the prevalence of EMRSA-15 amongst MRSA causing bacteraemia also reached its epidemic peak and was declining by the end of the first decade of this century. EMRSA-16 has also continued to decline since 2007 and this decline appears to have started naturally, prior to the overall MRSA bacteraemia decrease (Ellington *et al.*, 2010; Pearson *et al.*, 2009).

Figure 47. Trends in meticillin-resistant *S. aureus* (MRSA) bacteraemia from 1998 to 2009



The chart depicts trends in total proportion of *S. aureus* bacteraemia attributed to MRSA (blue line) and proportion of MRSA bacteraemia attributed to EMRSA-15 (red line), EMRSA-16 (green line) and other clones (purple line) in the UK. The dotted line represents proportions gained from the European Antimicrobial Resistance Surveillance System (EARSS) and the EMRSA-15 and EMRSA-16 strains were identified using bacteriophage typing (Johnson *et al.*, 2001). The solid line represents proportions gained from the BSAC and EMRSA-15 and EMRSA-16 were identified using PCRs which target R-M genes and SCC*mec* typing (Ellington *et al.*, 2010). The dashed line represents proportions identified in this study and EMRSA-15 and EMRSA-16 were identified using MLST and SCC*mec* typing.

This displays certain features of the perceived epidemic cycle where strains can undergo clonal expansion, followed by a relatively stable and successful period before they subsequently decline in the population (Wyllie *et al.*, 2011). It is unknown whether these

strains decline due to displacement by novel strains, environmental factors (infection control measures and antimicrobial therapy) and/or biological factors (host immune responses and the activity of bacteriophages) (Hibbing *et al.*, 2010; Rao *et al.*, 2011).

The case of EMRSA-15 and EMRSA-16 in the UK displays how the prevalence of epidemic MRSA clones can decline in the population. This decline may enable the opportunity for novel or less prevalent clones to emerge. Monitoring such trends could therefore provide insights into the success of certain *S. aureus* clones and enable us to determine the trends in prevalence of MRSA clones in the future. During the period 1998 to 2007, the proportion of MRSA bacteraemia isolates assigned to clones other than EMRSA-15 and -16 remained relatively constant. However, an increase was observed from 2007 to 2009 (Figure 47). This may indicate that the rapid decline of EMRSA-15 and -16 enables the population of low prevalent clones to increase as they are able to occupy niches previously dominated by these two strains. The CC22 and 30 lineages appear to be highly successful in the UK as they represented a large proportion of BSAC isolates (63 %). Although, sample bias may also account for the large proportion of CC22 and 30 isolates. Significant antimicrobial susceptibility differences were identified between CC22 and the other lineages which may explain, in part, the success of CC22 under conditions of selection due to antimicrobial therapy.

Low prevalence clones may represent rare or sporadic clones that have emerged recently or are less adapted to the hospital environment. These included ST5-MRSA-II, a clone thought to have a polyphyletic origin (multiple independent insertions of SCC*mec* type II) (Nübel *et al.*, 2008). MRSA clones also included ST45-MRSA-II which is rarely identified outside of the USA, the ST80-MRSA-IV clone which is often PVL-positive and an ST779 isolate which

is a sporadic lineage in the UK. The *mecA* gene but no recombinases were identified within the ST779 isolate and may indicate a combination of SCC*mec* and SCC*fus* elements (Monecke *et al.*, 2011). As the MRSA isolates assigned to clones other than EMRSA-15 and -16 belonged to individual MRSA lineages (CC5, 45, 80 and a singleton ST), this study did not provide evidence that these represented emerging MRSA clones. However, this study established the clones present at a lower frequency in 2009 which can be used as a baseline for future studies to monitor changes in the proportion of these clones in the UK.

MRSA isolates within the remaining collections in this study represented a diverse range of lineages. Although this study did not include representatives for the entire diversity of MRSA identified in the literature, the collection identified lineages which have given rise to numerous widespread HA and CA-MRSA strains through to sporadic strains (those that are identified at irregular intervals) in the UK. The few representatives of an ancestral MRSA clone (ST250-MRSA-I) indicated the continued state of decline of this clone, although the reasons for this remain unknown. SCC*mec* type III elements were restricted to ST239 isolates. These isolates also represented an ancestral clone which displays divergent hybridisation patterns as compared to other CC8 isolates as it was formed by a large-scale recombination event between isolates of CC8 and 30 (Monecke *et al.*, 2011; Robinson and Enright, 2004). This clone displays resistance to multiple antimicrobial classes and heavy metals which may explain its success in the UK and worldwide. Non-typeable SCC*mec* elements have been described in MRSA clones belonging to lineages such as CC8 (Monecke *et al.*, 2011). This may indicate the presence of composite and irregular SCC*mec* types that harbour novel virulence factors or antimicrobial resistance determinants that could play a role in the emergence of these clones. MRSA assigned to CC5 belonged to clones such as ST5-MRSA-II, the success of which may be driven by glycopeptide-resistant strains (Nübel *et*

al., 2008). CC45 MRSA included representatives of the sporadic CC45-MRSA-V clone and a ST45-MRSA-VI isolate, which to the best of the author's knowledge is a novel observation. This indicated the diversity amongst ST45 MRSA clones is increasing, although the reasons for this are unknown. Other clones such as ST1-MRSA-IVa, ST8-MRSA-IVa and ST97-MRSA-V have been associated with the arginine catabolic mobile element which may play a role in the virulence and/or fitness and hence the success of these clones (Diep *et al.*, 2008; Ellington *et al.*, 2008). MRSA isolates belonging to CC7, CC97 and ST361 are generally thought to be rare (Monecke *et al.*, 2011), whilst novel genotypes were identified amongst pandemic and sporadic lineages (CC7, 22, 45, 80 and 88). Therefore MRSA clones amongst the remaining collections included a range of diverse MRSA clones.

The CA-MRSA in this study were genetically diverse (CC1, 30, 59, 80, 88 and ST93) and have been associated with PVL (Monecke *et al.*, 2011). This may indicate these isolates are associated with severe disease (Gillet *et al.*, 2002). Despite being previously reported in the UK, clones assigned to ST93 are rarely identified in the UK. Monitoring the emergence of PVL-positive strains is vital as although the role of PVL in virulence is controversial, such strains are often able to infect individuals not associated with HA risk factors (Diep and Otto, 2008).

MSSA isolates were assigned to a greater number of lineages as compared to MRSA isolates and hence the isolates in this study represented a smaller proportion of the entire MSSA diversity than MRSA isolates described to date. The MSSA isolates belonged to diverse lineages which included those associated with major HA and CA-MRSA lineages, in addition to lineages that contain MSSA predominantly. Amongst MSSA, the greatest

proportion of isolates belonged to CC30 which correlates with studies indicating that this lineage is particularly successful in the UK (Ellington *et al.*, 2010; Johnson *et al.*, 2005).

Majority of the MSSA isolates were assigned to lineages similar to MRSA bacteraemia isolates (CC1, 8, 5, 22, 30, 45 and 59). This indicated that lineages have given rise to multiple successful MSSA and MRSA clones. Those isolates assigned to CC5 included an ST6 isolate which is a DLV of ST5. A large-scale chromosomal replacement between a CC5 isolate and a representative from another lineage is thought to have occurred to form ST6; approximately half of the ST6 genome is thought to originate from CC5. MSSA assigned to CC1 and 8 included an ST573 and ST72 isolate respectively, which based on microarray hybridisation patterns and MLST, indicates previous recombination events may have occurred to form these genotypes. The presence of multiple STs in these CCs demonstrated that indicated MSSA displayed considerable diversity amongst each lineage. The remaining MSSA isolates were composed of predominantly MSSA-specific lineages (CC9, 12, 15, 20, 25, 97, 101 and 121). CC15, for example, is a successful MSSA lineage which is a common colonising strain (Monecke *et al.*, 2011). The three remaining STs belonged to a lineage (ST425) associated with livestock, thus indicating the prevalence of infections associated with livestock-associated lineages remains rare at present. The presence of novel genotypes in MRSA-associated lineages (CC1, 5, 8, 30, 45 and 59), the predominantly MSSA CC121 and three singleton STs (ST2043, ST2044 and ST2221) identified in this study indicated these genotypes have emerged recently. Therefore this shows the diversity of MSSA continues to increase in a similar manner to that of MRSA. These findings are in accordance with previous findings which state MSSA isolates display greater genetic diversity than MRSA isolates. This is probably due to the limited SCC*mec* insertion events that have occurred in a

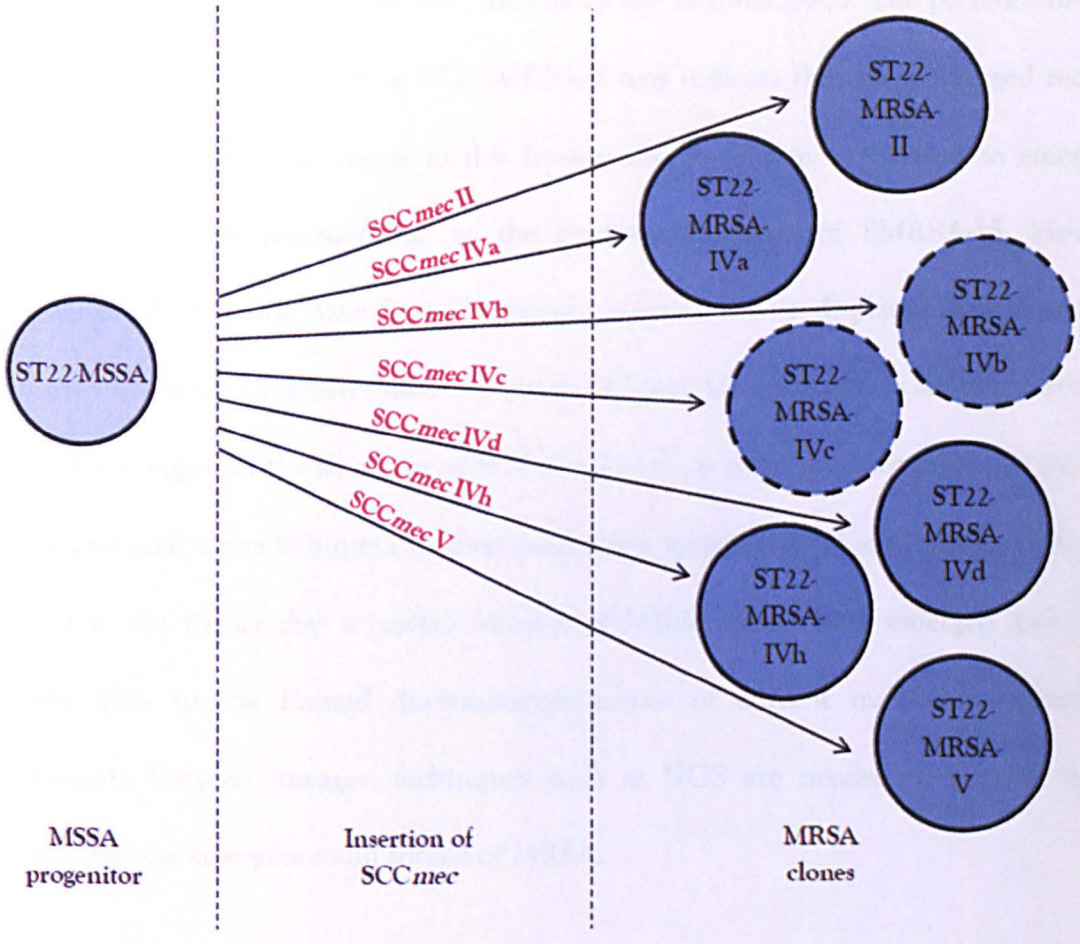
subset of genetic lineages to form a small number of MRSA clones which have spread globally (Fitzgerald *et al.*, 2001; Musser and Kapur, 1992; Robinson & Enright, 2003).

Five SCC*mec* types or subtypes and at least five non-typeable or non-subtypeable SCC*mec* elements identified in CC22 isolates indicated that this locus displayed considerable diversity within this lineage. Isolates with a SCC*mec* type IV non-subtypeable element revealed that these isolates may display variations in the virulence factors within the J regions of the element. Determining the sequence of these regions may be of value as it may identify the presence of SCC*mec* elements with novel virulence factors. SCC*mec* type IV and in particular, SCC*mec* IVh was overrepresented in CC22 which indicated EMRSA-15 was abundant and continued to be a significant cause of MRSA bacteraemia. Further investigation may identify whether this SCC*mec* subtype is directly associated with the prevalence of this strain in the UK. The ST22-MRSA-V clone identified amongst MRSA bacteraemia has been sporadically isolated in a few countries including the UK. Although the low prevalence of this clone indicates that it is not particularly successful, at present it is of value to monitor its emergence as PVL-positive versions of this clone have been identified (Boakes *et al.*, 2011; Monecke *et al.*, 2011). SCC*mec* non-typeable elements which consisted of *ccr* complex 4 and *mecA* indicated that SCC*mec* types VI and VII variants may be present in these isolates. Nevertheless, the presence of two *ccr* complexes in other non-typeable elements may indicate the presence of a composite SCC*mec*. However, further investigation would be required via determining the sequence of the SCC*mec* locus to confirm whether these findings indicate the presence of novel or variant SCC*mec* types which have not been previously described within this lineage. The diversity observed at this locus and within the housekeeping genes of the *S. aureus* MLST scheme demonstrates the concordance of this

study with previous studies that displayed divergence in this lineage across the genome as compared to other lineages (Monecke *et al.*, 2011).

A previous study has suggested the presence of multiple SCC*mec* IV subtypes amongst MRSA clones has arisen via independent SCC*mec* insertion events rather than conversion between subtypes via insertions or deletions. Therefore the presence of different SCC*mec* types and subtypes identified in the present and previous study indicated that insertion of different SCC*mec* cassettes has occurred on at least seven occasions (Boakes *et al.*, 2011; Mammina *et al.*, 2012). The identification of multiple ST22-MRSA-II isolates indicated one of these insertion events included the HGT of SCC*mec* type II into ST22, which to the best of the author's knowledge is a unique observation. Furthermore, as the replacement of an SCC*mec* type within an MRSA isolate is unfavourable, this indicates these insertion events have occurred into ST22-MSSA (Robinson & Enright, 2003). In some cases, the ST22-MSSA progenitor clones were PVL-positive which have given rise to PVL-positive MRSA clones. Further work is needed to identify whether ST22-MRSA-II encoded PVL. Nonetheless, these results have provided insights into the diversity and evolution of MRSA amongst CC22 (Figure 48).

Figure 48. Hypothetical insertions of SCC*mec* types and subtypes into ST22



The figure depicts seven hypothetical insertions into ST22-MSSA progenitors. In some cases, the progenitor clones were PVL-positive which gave rise to PVL-positive MRSA clones. MRSA clones illustrated in dashed circles were not identified in this study but have been identified in previous studies (Boakes *et al.*, 2011; Mammina *et al.*, 2012). MRSA clones in solid circles were represented in this study and ST22-MRSA-II is a novel clone identified only in this study.

The relationship of EMRSA-15 to the remaining ST22 isolates with SCC*mec* type II, V and other cassettes requires further investigation. If the insertion of alternative SCC*mec* types has occurred prior to the emergence of EMRSA-15 it poses the question as to why the ST22-

MRSA-IV combination is particularly successful. Conversely, insertion of alternative SCC*mec* types may have occurred after the emergence of EMRSA-15. The presence of novel or unusual MRSA clones such as ST22-MRSA-II may indicate they have emerged recently. Further SCC*mec* insertion events in this lineage may be driven or allowed to emerge by changes in selective pressure due to the continued decline of EMRSA-15. However, ST22-MRSA-II may not have been previously identified if it displayed low prevalence. Lineages such as CC22 which indicate multiple SCC*mec* insertion events support a previous study which suggested the insertion of SCC*mec* occurs on numerous occassions in the same lineage but each clone is limited in their geographic spread (Nübel *et al.*, 2008). This is in contrast to the theory that a limited number of MRSA clones have emerged and spread globally. Due to the limited discriminatory power of current molecular methods to discriminate between lineages, techniques such as NGS are needed to provide further insights into the emergence and spread of MRSA.

4.2. Genetic homogeneity between *S. aureus* lineages

This study identified twelve regions of genetic homogeneity amongst the MRSA lineages examined, of which a subset were mapped to WGSs. FAFLP identified the size of an AF, its sequence is unknown. Hence, we assume AFs of the same size originate from the same genomic locus. Thus, AFs present in all *S. aureus* isolates were taken as an indication that they originated from conserved regions of the genome which may encode important cell functions. The presence of AFs with a distinct sequence of the same size in the same isolate (collision) or in different isolates (homoplasy) may be a factor in the FAFLP profiles observed (Gort and van Eeuwijk, 2012).

Mapping one of the homogeneous regions to WGSs identified that it encoded the noncoding region between a recombination protein and the 16S rRNA gene. Despite not containing an ORF, this region may play a role in functions such as the regulation of neighbouring genes (Nübel *et al.*, 2008). The sequences of the remaining core regions were relatively conserved amongst lineages (98 to 100 %), and mapping to WGSs confirmed these regions harboured ORFs. Thus the presence of a specific AF in each lineage could identify the presence of highly conserved regions associated with essential cell functions. The presence of minor genetic variations agreed with previous findings which suggested variations within these regions were associated with significant phenotypic changes that affected strain growth and metabolism and may contribute to the survival, fitness and success of *S. aureus* (Table 37).

4.3. Genetic diversity between *S. aureus* lineages

This study utilised molecular methods to identify regions which display genetic variation which may confer a fitness advantage in major MRSA lineages by altering traits such as virulence or pathogenicity.

The presence of a unique combination of genetic variation in each lineage was confirmed by the clusters based on FAFLP data, which were concordant with MLST data, and the level of heterogeneity between lineages which was variable. The most successful lineage within the UK is CC22. The CC22 genome was highly divergent (≥47 % detected using FAFLP) from the remaining lineages, which is in accordance with previous studies (Monecke *et al.*, 2011). A proportionately high fraction of variations were indels (over 40 %) rather than SNPs. The majority of these indels were identified in regions encoding for lipoproteins or proteins

involved in the biosynthesis of amino acids. SNPs indicated regions which are associated with functions which may confer a selective advantage in this lineage and explain why this lineage predominates within the UK. CC22 isolates displayed distinct sequences for proteins which may alter the virulence and antimicrobial resistance. Additionally certain loci showed similarity with CC30 isolates, which may indicate cell growth, formation of the small-colony variant (SCV) phenotype and antimicrobial resistance may play a role in the success of these two lineages (Table 38).

The maximum heterogeneity was observed amongst CC8 as ST239 displayed considerable diversity from the remaining CC8 isolates, which is in accordance with previous findings (Monecke *et al.*, 2011). Alternatively, CC59 isolates displayed the greatest homogeneity, indicating few MRSA clones have emerged from this lineage. HA-MRSA lineages (CC5, 8, 22, 30 and 45) displayed greater divergence than CA-MRSA lineages (CC1 and 59). Further work on larger isolate sets would be required to determine whether this is due to differences in the relative ages of HA- versus CA- lineages or another factor such as differential antimicrobial selective pressure or their ability to spread in the nosocomial and community settings.

Based on FAFLP data, CC22 genome exhibited the greatest similarity with CC1, 5 and 8 genomes. CC1 and 5 genomes displayed the greatest homogeneity with each other. Genomes from other major lineages (CC30, 45, 59, and 97) displayed greater divergence from CC1, 5, 8 and 22. Based on loci throughout the genome, this study identified lineages which may be closely related to other lineages. Therefore further investigation of these loci may provide valuable insights into the evolution of MRSA lineages.

Table 37. Function of homogeneous regions amongst *S. aureus* lineages

Protein	Function	Phenotype of mutant	Reference
Transcriptional repressor (CodY)	Regulation of genes involved in nitrogen metabolism, amino acid biosynthesis, nucleotide metabolism and virulence	Display reduced rate of cell growth	(Majerczyk *et al.*, 2010; Pohl *et al.*, 2009)
Dihydroxyacetone kinase (DhaK)	Catabolism of glycerol an intermediate of glycolysis, the metabolic pathway which generates ATP	Lack of cell growth on glycerol	(Erni *et al.*, 2006; Pulsen *et al.*, 2000)
Phosphoglycerate kinase (PGK)	Role in glycolysis	Alters their growth under anaerobic conditions	(Nakano *et al.*, 1999; Watson *et al.*, 1982)
FtsK/SpoIIIE ATPase family	Segregation of chromosomes during cell division	Prevent cell division	(Begg *et al.*, 1995; Lyer *et al.*, 2004)
2-oxoisovalerate dehydrogenase	Catabolism of branched-chain amino acids	Unknown	(Aevarsson *et al.*, 1999)
Aminotransferase (NifS)	Mobility of sulphur for the formation of metalloenzymes	Unknown	(Zheng *et al.*, 1993)

Previous studies demonstrate the function of proteins identified in homogeneous regions amongst all *S. aureus* lineages and the effect of mutations within these protein coding regions.

WGM identified restriction fragment patterns specific to CC30, 45 and 59. Lineage-specific patterns may have been generated due to polymorphisms within regions harbouring the restriction endonuclease recognition sequence or insertion or deletion of acquired DNA elements. WGM alignments performed using ST8, ST30 and ST239 optical maps identified the regions of these genomes that were homogeneous between lineages. This supported the finding that ST239 was formed by a large recombination event (Robinson and Enright, 2003; 2004). The majority of the ST239 genome (76 %) was similar to the ST8 genome. Of

76 % of the ST239 genome, 74 % also displayed similarity to the ST30 genome, indicating that the remaining region (26 %) represents variable regions of the ST8 genome. The remaining 24 % of the ST239 genome was equally composed of ST30-specific regions or regions unique to ST239. The SCC*mec* region in ST239 did not show similarity with either ST8 or ST30 and hence this study confirms the origin of this element in ST239 remains unknown and requires further investigation (Holden *et al.*, 2010). Regions unique to ST239 indicated genetic variations or genetic events such as point mutations or HGT have arisen within these regions since the formation of ST239. Furthermore, unique fragment patterns were also identified within the former ST8 portion of the ST239 genome. This indicated the presence of genetic variations within these regions that have arisen after the recombination event between these lineages. Therefore genomic regions displaying lineage-specific heterogeneity require further investigation using NGS as they may provide insights into the survival and virulence of *S. aureus* lineages.

Where heterogeneous regions were identified between lineages, the regions were mapped to WGSs in order to identify whether these were attributed to differences in the distribution of these regions or whether they have been identified as a result of sequence variation. This study revealed the presence and absence of these AFs were due to sequence variation within the AF sequences in the form of SNPs, point mutations and small-scale indels at the loci identified. This is in accordance with a previous study which identified AFs specific to particular FAFLP clusters that are due to point mutations rather than the presence or absence of loci and these AFs are spread throughout the genome (Melles *et al.*, 2004). Regions identified in each lineage may encode essential cell functions and we expect these loci to be maintained via purifying selection, as minor variations in these sequences could result in significant phenotypic differences. As the majority of variations amongst these

sequences were specific to isolates of a specific lineage including MRSA and MSSA, this may indicate these variations arose prior to the insertion of SCC*mec* types in a common ancestor of each lineage. Conversely, these variations may have arisen via random mutations in a subset of the lineage and subsequently spread through the lineage via genetic transfer. Further analysis of these variations identified whether they altered the amino acid sequences of the encoded proteins which are now discussed below. The proteins included are those involved in functions which may provide a selective advantage.

The identification of lineage-specific variations could be utilised to develop assays which enable the rapid differentiation between lineages. Based on a single heterogeneous region, a pyrosequencing assay which could differentiate between major *S. aureus* lineages was developed. The assay was able to identify lineages for the majority of the isolates. Such assays would prove to be invaluable high-throughput and rapid tools, economical for the majority of resource-poor laboratories. However, further optimisation is required to improve the typeability of this assay. Further validation of the assay with a larger set of isolates may help to identify novel primer sequences. Such assays can help in outbreak situations to determine antimicrobial patterns and the genetic relatedness of the isolates. Furthermore, it can help us monitor trends in the emergence and spread of particular lineages.

A proportion of heterogeneous regions partly encoded proteins with a role in essential cell functions. These include proteins involved in aerobic and anaerobic respiration, biosynthesis of essential cell components, and DNA replication, repair and recombination. Other regions encoded proteins involved in nonessential functions such as antimicrobial resistance, virulence and pathogenicity (Table 38). Previous studies have revealed sequence variations within these regions can influence the growth, virulence, pathogenicity or antimicrobial

resistance of *S. aureus*. Specific regions were exclusive to isolates of one lineage or a subset of lineages indicating they represented accessory genomic regions. However, in the majority of cases, heterogeneous regions were identified in each lineage.

The majority of variations associated with regions encoding respiratory functions were restricted to non-coding regions, which can affect gene regulation (Gottesman, 2005). On the contrary, those encoded on ORFs were synonymous, indicating the conservation of protein sequence and function between lineages, in the main. The AF sequence for F47, nitrite reductase involved in the anaerobic conversion of nitrite to ammonia, was conserved amongst the majority of lineages but displayed numerous CC22-specific SNPs of which most were non-synonymous. With regards to the mutations in the non-coding regions of the AF, considering the effect of mutations in the regulatory systems which control the transcription of the nitrite operon is useful. Mutations in these loci result in decreased nitrite reductase activity and formation of a SCV phenotype under anaerobic conditions (Schlag *et al.*, 2008). SCVs show an increase in antimicrobial resistance and persistence in the host (Tuchscherr *et al.*, 2011). Further investigation would be required to identify whether the non-synonymous CC22-specific SNPs alter the function of nitrite reductase, and /or whether the synonymous SNPs alter the translation efficiency of the protein. If either hypothesis (above) is correct, then this region may alter the growth and survival of CC22 isolates under anaerobic conditions.

Haem is the prosthetic group of a number of proteins including the cytochromes involved in the respiratory electron transport chain (Lee, 1995; Tien & White, 1968). Multiple heterogeneous regions (F38, F45 and F49) which encoded proteins involved in haem biosynthesis were identified in this study. These proteins were conserved amongst the

majority of lineages except two major lineages CC8 and 22, in addition to ST93 where variations occurred. Mutations in proteins with a role in haem biosynthesis can result in a SCV phenotype which can also be associated with repressed levels of virulence factors such as α-haemolysins (Kohler *et al.*, 2003). Such factors may aid in evasion of the host immune system and variations within this region may affect the growth and virulence. Another protein (F21) involved in the biosynthesis of cobalamin was specific to CC22. Cobalamin is a coenzyme required for multiple metabolic processes, and the biosynthesis of cobalamin can occur via multiple pathways under aerobic and anerobic conditions (Raux *et al.*, 2000). Further work is required to establish the function of cobalamin in providing a selective advantage for this succesful lineage and to determine whether variations in haem biosynthesis associated regions aid the evasion of the host immune response in CC8, 22 and ST93.

Proteases were encoded on multiple heterogeneous regions, one of which included dipeptidyl aminopeptidase (F40) which cleaves the terminus of peptides with an alanine or proline as the penultimate amino acid. Aminopeptidases have a role in virulence, nitrogen metabolism and the activation or degradation of various proteins (Gonzales and Robert-Baudouy, 1996). The disruption of genes encoding for aminopeptidases have been shown to reduce the severity of an *S. aureus* infection (Carroll *et al.*, 2012). The greater proportion of SNPs specific to CC22 and 59 were synonymous whilst those specific to CC5 and 30 were non-synonymous. This indicated the function of this protein is conserved in certain lineages such as CC22 to a greater extent than in others such as CC30. Cell functions such as replication, repair and recombination of DNA were encoded on multiple heterogeneous regions. SNPs at these loci were restricted to non-coding regions or were synonymous, indicating the essentiality and conservation of the amino acid sequence and

may indicate that these loci are under the action of purifying selection. The non-coding region displayed sequence divergence particularly in CC30 and 45 isolates as compared to other lineages and further investigation is required to identify whether these SNPs alter the regulation of DNA repair and whether this affects strain success.

One heterogeneous region encoded a protein (F13) which interferes with the development of bacteriophages after their DNA has penetrated the cell (Duckworth *et al.*, 1981). This mechanism often results in cell death as the infection has progressed sufficiently to damage cell functions or host DNA (Garvey *et al.*, 1995). The one non-synonymous SNP identified within this region was specific to CC8. Whilst further work is necessary to investigate the significance of the SNP, mutations within proteins displaying abortive infectivity domains are known to affect the quantity of protein A and a subset of other surface proteins. In addition, this alters the abundance and shape of the cross-walls forming the peptidoglycan. Therefore mutations in these proteins may affect cell division, cell separation and virulence (Frankel *et al.*, 2010).

A number of proteins encoded on heterogeneous regions play a role in the efflux or degradation of antimicrobials. A partial ABC transporter protein (F23) required for the efflux of multiple antimicrobials was identified in CC22, 30, 59 and ST425. Previous studies have identified mutations in these proteins result in an increased sensitivity to the antimicrobial bacitracin (Ohki *et al.*, 2003). Multiple regions encoded lipoproteins, one of which (F22) was present in the majority of lineages excluding CC1, 30 and 59. This protein displayed a unique sequence in CC5, 8, 22 and ST93 isolates. Lipoproteins play numerous roles in *S. aureus* which include immune evasion and antimicrobial resistance. The absence of this protein in certain CA-MRSA lineages and the unique protein sequence displayed

amongst other lineages may explain, in part, differences in antimicrobial resistance or virulence between lineages. Two lipoproteins displayed a divergent amino acid sequence in CC30 (F19) and CC5 (F48) respectively. Post-translational modification of lipoproteins has been shown to play an important role in pathogen recognition via the innate immune system (Wardenburg *et al.*, 2006). Lipoproteins are also essential to maintain intermediate glycopeptide resistance (Jousselin *et al.*, 2012). Variations amongst these proteins could alter the ability of this organism to resist antimicrobial therapy and cause infection. It appears that these regions may be under greater selective pressure in CC5 and 30 as compared to other lineages. Further to antimicrobial resistance, herbicidal resistance could affect environmental resistance or persistence for *S. aureus*. Herbicides may affect bacterial growth and hence genes which confer resistance to herbicides may provide a selective advantage where herbicides are used. An enzyme involved in the degradation of herbicides (F15; allophanate hydrolase) was identified in one region (Cheng *et al.*, 2005). The majority of SNPs in this region specific to CC22 and 30 were synonymous whilst those specific to CC8, 59 and ST93 were non-synonymous. These proteins are found throughout bacteria and display a region of conserved residues of up to 130 amino acids. Mutations within the codons of these amino acids have not been documented in *S. aureus*. However, in *Granulibacter bethesdensis*, a Gram-positive environmental bacterium, mutations within this protein severely reduce the catalytic activity of the protein (Lin and St Maurice, 2013).

Variation of the protein A sequence was concordant with MLST, indicating this region displayed lineage-specific variation. Furthermore, the presence of the *spa* gene in all *S. aureus* lineages indicated it is essential for the survival of *S. aureus*. Lineage-specific variation has been identified in the *spa* gene and certain domains of the protein A, particularly those at the host-interface, display a higher level of interlineage variation compared to other domains

(McCarthy and Lindsay, 2010). VNTRs are thought to aid the organism to adapt to environmental factors by accumulating variation providing a selective advantage (van Belkum *et al.*, 1998). Therefore, conserved regions of the *spa* gene amongst each lineage indicates that each lineage may have adapted differently to evade the host immune response (McCarthy and Lindsay, 2010). Therefore further studies may indicate whether this variation plays a role in the success of a particular lineage.

Further acquired regions for which variable AFs were found included F10, which partially encoded proteins with homology to an addiction toxin/antitoxin module. This particular system showed homology to the Txe toxin from *Enterococcus faecium*, which when the toxin is bound to the antitoxin, leaves the cell unaffected. However, the toxin is freed once the antitoxin is degraded, resulting in cell damage under conditions of selection due to nutritional or environmental stress; moreover this system has been shown to affect cell growth and division (Engelberg-Kulka and Glaser, 1999; Grady and Hayes, 2003). The majority of SNPs amongst this sequence specific to CC1 and 59 were synonymous whilst the majority specific to CC5, 22 and 30 were non-synonymous. Further analysis would be required to confirm whether CC5, 22 and 30 isolates encode a functional addiction module and whether these variations cause differences in the ability of each lineage to adapt or survive in certain environments. A further fragment (F29) from isolates in CC30 also mapped to a region in ST239 indicating this region represents a region that recombined from ST30 to form ST239 (Robinson and Enright, 2003; 2004) and may represent the SCC*mec* chromosomal junction, although this would require further analysis.

Table 38. Variations in function of proteins between lineages

KEGG protein functional group	Specific function	Clonal complex									Effect of mutations	References
		CC1	CC5	CC8	CC22	CC30	CC45	CC59	CC97	ST93		
Carbohydrate metabolism	Conversion of pyruvate to acetate or lactate and virulence factor	N	Y	N	N	N	N	N	N	N	Reduced enzyme activity, increase in pH, reduced virulence	(Patton *et al.*, 2005; Richardson *et al.*, 2008)
Nitrogen metabolism	Conversion of nitrate to ammonia	N	N	N	Y	Y	N	N	N	N	Decreased nitrite reductase activity, formation of a small-colony variant phenotype	(Schlag *et al.*, 2008; Tuchscherr *et al.*, 2011)
Metabolism of cofactors and vitamins	Synthesis of the prosthetic group of a no. of proteins	N	N	Y	Y	N	N	N	N	Y	Formation of small-colony variant phenotype and exhibition of repressed levels of virulence factors	(Kohler *et al.*, 2003)
Biosynthesis of amino acids	Biosynthesis of branched-chain amino acids	Y	Y	Y	Y	Y	Y	Y	Y	Y	Lack of growth when acetate is the only carbon source	(Dailey & Cronan, 1986)
DNA replication and repair	Repair of damaged DNA.	N	N	N	N	Y	N	N	N	Y	Reduced rate and efficiency of DNA repair	(Honda *et al.*, 2008; Inoue *et al.*, 2008; Liu *et al.*, 2011; Marceau, 2012;)

Table 38 continued. Variations in function of proteins between lineages

KEGG protein functional group	Specific function	Clonal complex									Effect of mutations	References
		CC1	CC5	CC8	CC22	CC30	CC45	CC59	CC97	ST93		
Cell division	Unlink and segregate chromosomes	N	N	N	N	N	N	N	N	N	Forms multi-chambered chains	(Begg *et al.*, 1995; Geissler and Margolin, 2005)
Proteolysis	Cleaves the terminus of certain peptides	N	Y	Y	Y	Y	Y	Y	N	Y	Reduced enzyme activity and virulence	(Kumagai *et al.*, 2000)
Bacteriophage resistance	Interferes with the development of bacteriophages	N	N	Y	N	N	N	N	N	N	Effects cell division, cell separation and virulence	(Frankel *et al.*, 2010)
Signal transduction	Regulation of cell wall growth, initiation of DNA replication and efflux of bacitracin. Transfer of a methyl group to rRNA.	Y	N	N	N	N	N	Y	N	N	Increased sensitivity to bacitracin and penicillin, altered resistance to multiple antimicrobials and no DNA replication at higher temperatures	(Butler *et al.*, 2002; Dintner *et al.*, 2011; Groicher *et al.*, 2000; Inoue *et al.*, 2001; Kawada-Matsuo *et al.*, 2011; Long *et al.*, 2006; Ohki *et al.*, 2003)

Table 38 continued. Variations in function of proteins between lineages

KEGG protein functional group	Specific function	Clonal complex									Effect of mutations	References
		CC1	CC5	CC8	CC22	CC30	CC45	CC59	CC97	ST93		
Drug resistance	Active efflux or degradation of multiple antimicrobial molecules.	Y	Y	Y	Y	Y	N	Y	Y	Y	Increased meticillin resistance, enables evasion of an immune response and altered oxacillin and glycopeptide resistance	(Jousselin *et al.*, 2012; Lin and St Maurice, 2013; Sieradzki *et al.*, 2008; Wardenburg *et al.*, 2006)
Genome stability factors	Factors involved in integration or loss of DNA	Y	Y	Y	Y	Y	N	Y	N	Y	Lack of formation of small capsids	(Damle *et al.*, 2012; Poliakov *et al.*, 2008)

Displays KEGG functional group (obtained from UniProt database; http://www.uniprot.org) and function of proteins encoded in regions which displayed lineage-specific variation. Displays whether protein function is altered (Y) or conserved (N) in each lineage and the effect variations within these proteins may have in each lineage based on previous studies (see references).

Sequences which encoded proteins that play a role in essential cell functions such as haem biosynthesis, DNA repair and cell division displayed a greater number of synonymous SNPs. Proteins which were associated with resistance of the organism to certain antimicrobials or drugs also displayed a bias of synonymous SNPs. Alternatively, proteins such as transposases, lipoproteins, bacteriophage-related proteins or those with homology to those with a role in apoptosis in eukaryotes displayed more non-synonymous SNPs. However, most AF sequences only partially encoded an ORF, it was difficult to deduce whether purifying selection may be acting on particular regions.

The number of synonymous versus non-synonymous mutations can indicate the level of amino acid sequence conservation between lineages. However, as dN/dS ratios could not be calculated, the number of synonymous and non-synonymous SNPs could not be taken as a definite indicator of selective pressure within these regions. Synonymous mutations can alter the translation accuracy and efficiency of the protein which may, in turn, affect the function of the protein. Where non-synonymous mutations occur, selection for codon bias is likely to occur as certain codons are translated more accurately or efficiently (Hershberg and Petrov, 2008). Non-synonymous mutations result in an amino acid change, the effect of this substitution is dependent on numerous factors. If the biochemical properties of the substituted amino acid are similar to the original, the effect of this substitution may be minimal. Also the position of the amino acid in relation to the protein structure is significant as changes to particular residues in the active site of proteins may have a profound effect on the function of the protein as compared to those in other regions. This study identified regions of variation which may alter the regulation or function of proteins between lineages. However, further work would be needed to confirm whether differences in the expression or function of these proteins could provide a selective advantage in certain

MRSA lineages which contribute to their success.

4.4. Genetic diversity within *S. aureus* lineages

Current molecular methods such as MLST and *spa* typing were able to identify the heterogeneity between lineages. However, limited genetic variation was identified within lineages. Hence techniques which are able to identify the heterogeneity within lineages are needed. Although in outbreak situations the relatedness of isolates may be indicated with current methods, a limited number of lineages often circulate in a given hospital. Therefore methods which can distinguish between strains such as EMRSA-15 and the remaining CC22 isolates are required. This could help us identify whether predominant MRSA strains represent a distinct subset of these lineages. From this information we may be able to identify regions which play a role in the success of these strains.

The SCC*mec* types or subtypes identified in each lineage indicated the minimum number of insertions which have occurred in each lineage. As SCC*mec* types are rarely replaced by another, this indicated each SCC*mec* insertion is likely to have occurred into an MSSA representative of each lineage (Robinson & Enright, 2003). Major HA-MRSA lineages such as CC5, 8, 22 and 45 displayed five or more SCC*mec* types/subtypes whilst CA-MRSA such as CC1, 59, 80 and ST93 displayed up to three SCC*mec* types/subtypes. This indicated HA-MRSA displayed greater genetic heterogeneity at this locus as compared to CA-MRSA. The SCC*mec* region harbours multiple antimicrobial resistance determinants; this variation may reflect how this pathogen has adapted to the greater antimicrobial selective pressure in hospitals as compared to the community. However, a greater number of isolates represented HA-MRSA lineages as compared to CA-MRSA lineages. Hence further investigation is

required to identify whether a significant difference is present between HA and CA-MRSA. Similar to the SCC*mec* locus, HA-MRSA lineages displayed greater heterogeneity at the *spa* locus as compared to CA-MRSA. Variations in VNTRs can affect virulence (van Belkum *et al.*, 1998). Therefore further investigation of differences within lineages at this locus may identify whether lineages have adapted differentially to the host. Novel *spa* types identified amongst CC22 and 30 isolates indicated these lineages are continuing to evolve at this locus. As multiple *spa* types were identified in the same lineage, this also indicated that this locus evolves at a different rate to that of the core genome (van Belkum *et al.*, 1998). These data may help ascertain whether certain lineages are under greater selective pressure at this locus and whether these variations alter the virulence of these lineages.

4.4.1. Genetic diversity within lineages across the genome

The heterogeneity within lineages was identified using FAFLP. Subsets of isolates in each lineage displayed greater similarity to each other than the remaining isolates of that lineage, indicating these isolates were closely related and may have arisen from a recent common ancestor. The majority of isolates displayed unique FAFLP profiles which indicated heterogeneous regions unique to a single isolate. This heterogeneity may have arisen via point mutations. However, as these variations were restricted to a particular lineage, this may indicate these variations have arisen recently or that these variations may not be beneficial and hence are not under the influence of natural selection. Heterogeneous regions unique to a subset of isolates of one lineage were identified in CC1, 8, 22 and 80. This indicated these regions displayed variation which was either present in a recent common ancestor of these isolates or this variation has emerged and spread amongst a subset of the lineage. Further investigation of these regions may identify whether within CC22, for example, these heterogeneous regions were unique to EMRSA-15 and whether they may play a role in the

predominance of this strain.

The level of heterogeneity identified within lineages was variable. Despite the fact that CC22 isolates were overrepresented in this study these isolates displayed the greatest homogeneity. This indicates the majority of MRSA isolates assigned to CC22 have emerged from clones formed from a relatively few number of SCC*mec* insertion events. CC22 isolates formed two distinct genetic groups; one consisted of the majority of isolates from CMPHL whilst the remaining cluster grouped CC22 isolates from the remaining collections. Therefore isolates from the same lineage were genetically diverse, based on their geographic location which indicated these groups are likely to have arisen via independent SCC*mec* insertion events. Further work is therefore needed to identify whether regions of genetic heterogeneity between these two clusters provide insights into how isolates adapt differentially to their environment. CC8 isolates formed two clusters; one had the majority of MSSA isolates whilst the remaining cluster predominantly had MRSA isolates. This indicated CC8 MRSA display considerable divergence from MSSA isolates of the same lineage. The remaining lineages formed single genetic groups composed of MRSA and MSSA isolates, thus indicating MRSA isolates are genetically similar to MSSA isolates of the same lineage. Each lineage displayed a variable level of heterogeneity; however as the number of isolates assigned to each lineage was variable, further work would be needed to confirm which lineages displayed greater genetic diversity as compared to other lineages.

Restriction fragment patterns unique to a subset of isolates amongst those assigned to ST8, ST22 and ST239 were identified, thus indicating the presence of variation specific to those isolates. Mapping of these regions to WGSs indicated that these regions may harbour surface proteins such as SasH which exhibit intralineage variation. Furthermore, MGEs such as

SCC*mec*, bacteriophage proteins and the *blaZ* operon, which show transposition activity *in vitro*, were identified in these regions indicating HGT of this MGEs may have occurred into ST239 (Griffin *et al.*, 1999; McCarthy and Lindsay, 2010). Variations within such regions may alter the virulence or antimicrobial resistance of these subsets of isolates contributing to their success (Butler *et al.*, 2002;Richardson *et al.*, 2008).

Restriction fragment patterns also identified heterogeneity within CC22 isolates. With the exception of NGS, more than one molecular method is often required to identify regions of genetic diversity within this lineage (Shore *et al.*, 2010). One region of the CC22 genome revealed an isolate-specific inversion. Part of the restriction fragment pattern exhibited within the inverted region was repeated adjacent to the inverted region. The repeated region encoded genes for surface-associated proteins such as protein A, capsular polysaccharide synthesis protein and lipoproteins. This inversion also harboured two transposases, thus indicating that the inverted region was formed by a transposition event. However, this inversion may also be the result of site-specific recombination or double-strand break repair. The stability of inversions and their effect on cell fitness is variable (Campo *et al.*, 2004). Furthermore, inversions can alter the expression of genes (Henderson *et al.*, 1999). Therefore, further studies would provide an insight into the effect of this inversion on the virulence and cell fitness of this isolate.

The presence of restriction fragment patterns unique to a subset of isolates within a lineage indicate the presence of intralineage variation. Surface proteins within such regions have been shown to display intra-lineage variation. The level of variation is greater amongst lineages than within lineages which shows that these proteins are relatively conserved within a lineage. However, levels of intralineage variation differed between lineages which may

correspond to differences within each lineage to adapt to the host or evade an immune response (McCarthy and Lindsay, 2010). Techniques such as MLST are limited in their ability to identify differences between STs. Therefore tools such as FAFLP and WGM can help identify regions of the genome which display variation within lineages. Further investigation by determining the sequence of these regions could identify which proteins display intralineage variation and whether the function of these proteins differ between variants. This may help us identify differences which contribute to the success of epidemic strains such as EMRSA-15.

4.5. Conclusions

This study investigated the genetic diversity amongst predominant HA-MRSA lineages. Distinct genetic differences were identified between successful and unsuccessful lineages of MRSA. Regions of genetic variation were identified between and within lineages. Greater diversity was exhibited between than within these lineages. In part, genetic differences between lineages were attributed to regions distributed exclusively amongst a particular lineage. Regions unique to each of three CCs (CC8, 22 and 30) were identified, which may alter the virulence or antimicrobial resistance amongst each lineage and contribute to their success. The majority of the heterogeneity observed within the cultures tested, however, was due to lineage-specific sequence variation at loci throughout the genome. The proteins encoded at these loci indicated they play a role in the emergence and prevalence (or spread) of dominant genetic lineages by altering features such as virulence or antimicrobial resistance.

The CC22 genome was highly divergent from the remaining lineages and the presence of novel genotypes indicates the level of this diversity continues to increase. Regions of

heterogeneity were associated with virulence and antimicrobial resistance which may confer a selective advantage in this lineage and explain why this lineage predominates within the UK. Additionally certain loci showed similarity with CC30 isolates, which indicated cell growth, formation of the SCV phenotype and antimicrobial resistance may play a role in the success of these two lineages.

Factors with a role in carbohydrate metabolism are unlikely to contribute to the success of any particular lineage as these proteins appear conserved between lineages. However, proteins which are involved in aerobic or anaerobic respiration may reduce the virulence exhibited by CC5. The virulence of CC22 and 30 appear to be altered due to variations identified in proteins with a role in nitrogen metabolism, which may explain in part the success of these lineages in the UK. Heterogeneity amongst proteins with a role in the metabolism of cofactors and vitamins may contribute to changes in the virulence of CC8, 22 and ST93. The predominance of multiple lineages may be attributed to variations in proteins involved in the biosynthesis of amino acids, as such variations may hamper the ability of isolates of certain lineages to grow and survive. CC30 and ST93 isolates displayed variations amongst proteins with a role in DNA repair. Therefore variations in these regions may alter the ability of the cell to repair damaged DNA which may provide a selective disadvantage amongst these lineages. Proteins required for essential cell functions such as cell division are unlikely to contribute to the success of any particular lineage as these proteins are relatively conserved between lineages. The diversity amongst proteases was identified to alter virulence and play a role in the success of multiple lineages. The heterogeneity amongst bacteriophage resistance proteins was also identified to contribute to the success of CC8 by causing changes in virulence in this lineage. The predominance of a subset of lineages may be attributed in part to diversity amongst proteins with a role in signal

transduction or drug resistance; heterogeneity within these regions may alter the resistance of these lineages to multiple antimicrobials, thus aiding their survival under conditions of selection due to antimicrobial pressure. This study also identified the success of multiple lineages is partially attributed to heterogeneity within proteins with a role in genome stability as this may affect genomic transfer in these lineages. Therefore it appears as though no single factor contributes to the success of an MRSA lineage. Multiple contributing factors which alter features such as virulence, antimicrobial resistance and HGT play an important role in the success of predominant MRSA lineages.

The ability of a pathogen to cause infections depends on multiple interactions between the host and pathogen. Therefore a combination of host and pathogen factors is important to cause disease. However, this study focused only on factors relating to the predominance of particular lineages encoded on the bacterial genome. The genome of the host has been shown to be associated with the genotype of *S. aureus*. Therefore factors relating to the predominance of particular lineages may in part be related to host factors (Emonts *et al.*, 2008). Furthermore, although the bacteraemia collection provided a well characterised and representative isolate collection, these results are unable to provide many insights into isolates associated with carriage or other types of infection. Particularly as the predominant *S. aureus* genotypes may be variable depending on the site of infection or isolates associated with different infections may exhibit different phenotypes such as antimicrobial resistance (Reynolds, 2009). Although the findings from this study may not be extrapolated to isolates from other infection sites, a similar approach could be used to analyse such isolates in the future.

Various molecular methodologies in this study have been demonstrated as useful tools to

investigate the genetic diversity between and within MRSA lineages. In addition to identifying regions of genetic homogeneity these tools enabled the detection of heterogeneity between lineages. It was shown that FAFLP analysed conserved regions in a similar manner to MLST and could be utilised to infer the macro-evolution of MRSA lineages (Lindsay & Holden, 2004). Moreover, whilst FAFLP probably underestimated the level of genetic diversity in accessory regions, it sampled variable regions thus facilitating understanding of the micro-evolution of *S. aureus* and the diversity of MGEs amongst lineages (Melles *et al.*, 2004; van Leeuwen *et al.*, 2005).

4.6. Future direction

A subset of proteins encoded in heterogeneous regions remained unidentified. Further work such as gene knockout studies and transcriptional studies using microarrays would be needed to identify the functions of these proteins and the effect of the amino acid changes to its function and phenotype of the bacterium. This study identified regions unique to a subset of lineages which may alter the regulation or function of proteins between lineages. However, further work would be needed to confirm whether differences in protein function could provide a selective advantage to certain MRSA lineages which contribute to their success. Furthermore, regions identified in a limited number of lineages are found to be associated with MGEs such as the SCC*mec*. MGEs may be restricted to certain lineages by mechanisms such as R-M systems which limit the exchange of genetic information between lineages. Further work would be needed to identify whether a unique combination of MGEs, which may harbour antimicrobial resistance determinants or virulence factors, are present in each lineage. The diversity of MGEs may explain in part the success of particular lineages under certain conditions such as antimicrobial therapy. The acquired SCC*mec* type

IV element was found in almost every lineage, indicating that combinations of this SCC*mec* type and the MRSA lineages were stable and often gave rise to successful MRSA clones. However, SCC*mec* type IV is highly variable with at least eight sub-types known to exist. The overall success of SCC*mec* type IV and the diversity of sub-types are thought to be associated with a lower fitness cost and higher mobility for the relatively small SCC*mec* type IV element (Ma *et al.*, 2002).

Existing molecular typing techniques for MRSA may not detect minor variation throughout the genome. NGS provides the ultimate resolution for detecting genetic diversity as it can distinguish between isolates with variations at a single nucleotide position. NGS has been useful for the development of microarrays which have been utilised for comparative genomics, identification of new genes and functions in *S. aureus*, and has been applied in outbreak situations. Since the advent of NGS, the cost and time to sequence a genome have fallen dramatically. These improvements could lead to NGS being utilised in laboratories as a routine procedure in the near future. At present however, it is not yet feasible for the majority of laboratories to resource the technique.

NGS can be utilised to monitor the emergence, spread and evolution of clones (Köser *et al.*, 2012). CC22 has been shown to acquire a wide-range of MGEs associated with antimicrobial resistance; however, these MGEs are not accumulated. NGS could aid the monitoring of the acquisition and loss of such elements which are thought to be the key to the success of this lineage (Lindsay *et al.*, 2012). The greatest hurdle to introduce NGS into routine practice is the lack of bioinformatics tools required to analyse the large volume of NGS data generated. The development of databases for comparisons and an automated analysis method are required to identify rapidly whether isolates are associated with an outbreak and how the

emergence of variations play a role in the evolution of clones.

NGS can also be useful in systems biology to model complex biological networks and pathways which could be utilised to determine the effect of individual genes on the bacterium (Holden *et al.*, 2008). Such analysis requires a complete view of the pan-genome of *S. aureus*. The pan-genome consists of all genes identified in members of a species. Each whole-genome sequenced strain identifies novel genes which continuously expands our knowledge of the *S. aureus* pan-genome which indicates this pan-genome is affectively open-ended (Holden *et al.*, 2008; Mira *et al.*, 2010). In this situation, techniques such as FAFLP may provide a feasible alternative to study the pan-genome until the cost and analysis of WGSs are further improved. FAFLP could be utilised to identify rapidly regions of heterogeneity which have not been previously described enabling the sequence determination of regions of interest or WGSs of the subset of isolates which display heterogeneity.

Bacterial cells have complex systems of biochemical pathways. Each gene is not necessarily transcribed into a protein and each protein does not necessarily have a single function. Hence from this study, it is clear further work is needed to identify whether heterogeneity is present across the entire protein coding sequences of interest. Future studies would require gene knockout studies of these regions to confirm the effect of this heterogeneity on the function of these proteins. Previous studies have shown resistance to certain antimicrobials contribute to the success of certain lineages. Lineages such as CC22 have been shown to display significant diversity of MGEs which is thought to contribute to the fitness of this clone (Lindsay *et al.*, 2012). Therefore further work would also be required to identify whether changes in the functions of interest identified in this study alter the fitness of

particular lineages. Understanding the genetic diversity amongst lineages can provide insights into the population genetics of MRSA and the evolution of successful lineages. This study indicates a combination of factors play a role in the success of major MRSA lineages. The genetic diversity amongst core regions and MGEs appear to alter antimicrobial resistance and virulence factors which are vital to their success as they enable the rapid adaption of lineages to their environment.

Bibliography

Abraham E, Chain E. An enzyme from bacteria able to destroy penicillin. Nature. 1940;146(3713): 837.

Aevarsson A, Seger K, Turley S, Sokatch J, Hol W. Crystal structure of 2-oxoisovalerate and dehydrogenase and the architecture of 2-oxo acid dehydrogenase multienzyme complexes. Nature Structural Biology. 1999; 6(8): 785-792.

Ahmadian A, Gharizadeh B, Gustafsson A, Sterky F, Nyrén P, Uhlén M, Lundeberg J. Single-nucleotide polymorphism analysis by pyrosequencing. Analytical Biochemistry. 2000; 280(1): 103-110.

Allen H, Donato J, Wang H, Cloud-Hansen K, Davies J, Handelsman J. Call of the wild: antibiotic resistance genes in natural environments. Nature Reviews Microbiology. 2010; 8(4): 251-259.

Andrews J. Determination of minimum inhibitory concentrations. Journal of Antimicrobial Chemotherapy. 2001; 48(Supplement S1): 5-16.

Anonymous. The Science Creative Quarterly. [Online]; 2003 [cited 2013]. Available from: http://www.scq.ubc.ca/genome-projects-uncovering-the-blueprints-of-biology/.

Antonio M, McFerran N, Pallen M. Mutations affecting the rossman fold of isoleucyl-tRNA synthetase are correlated with low-level mupirocin resistance in *Staphylococcus aureus*. Antimicrobial Agents and Chemotherapy. 2002; 46(2): 438-442.

Archer G, Thanassi J, Niemeyer D, Pucci M. Characterization of IS*1272*, an insertion sequence-like element from *Staphylococcus haemolyticus*. Antimicrobial Agents and Chemotherapy. 1996; 40(4): 924-929.

Armstrong-Esther C, Smith J. Carriage patterns of *Staphylococcus aureus* in a healthy non-hospital population of adults and children. Annals of Human Biology. 1976; 3(221-227): 3.

Arnold C, Metherell L, Willshaw G, Maggs A, Stanley J. Predictive fluorescent amplified-fragment length polymorphism analysis of *Escherichia coli*: high-resolution typing method with phylogenetic significance. Journal of Clinical Microbiology. 1999; 37(5): 1274-1279.

Ashelford K, Chuzhanova N, Fry J, Jones A, Weightman A. At least 1 in 20 16S rRNA sequence records currently held in public repositories in estimated to contain substantial anomalies. Applied and Environmental Microbiology. 2005; 71(12): 7724-7736.

Aucken H, Ganner M, Murchan S, Cookson B, Johnson A. A new UK strain of epidemic methicillin-resistant *Staphylococcus aureus* (EMRSA-17) resistant to multiple antibiotics. Journal of Antimicrobial Chemotherapy. 2002; 50: 171-175.

Baba T, Bae T, Schneewind O, Takeuchi F, Hiramatsu K. Genome sequence of *Staphylococcus aureus* strain Newman and comparative analysis of staphylococcal genomes: polymorphism and evolution of two major pathogenicity islands. Journal of Bacteriology. 2008; 190(1): 300-310.

Baba T, Takeuchi F, Kuroda M, Yuzawa H, Aoki K, Oguchi A, Nagai Y, Iwama N, Asano K, Naimi T, Kuroda H, Cui L, Yamamoto K, Hiramatsu K. Genome and virulence determinants of high virulence community-acquired MRSA. The Lancet. 2002; 359(9320): 1819-1827.

Bannerman T, Hancock G, Tenover F, Miller J. Pulsed-field gel electrophoresis as a replacement for bacteriophage typing of *Staphylococcus aureus*. Journal of Clinical Microbiology. 1995; 33(3): 551-555.

Bayer A, Sullam P, Ramos M, Li C, Cheung A, Yeaman M. *Staphylococcus aureus* induces platelet aggregation via a fibrinogen-dependent mechanism which is independent of principal platelet glycoprotein IIb/IIIa fibrinogen-binding domains. Infection and Immunity. 1995; 63(9): 3634-3641.

Beck W, Berger-Bächi B, Kayser F. Additional DNA in methicillin-resistant *Staphylococcus aureus* and molecular cloning of *mec*-specific DNA. Journal of Bacteriology. 1986; 165(2): 373-378.

Becker W, Kleinsmith L, Hardin J. Sexual reproduction, meiosis, and genetic recombination. In The world of the cell. 6th ed. San-Francisco: Benjamin Cummings; 2006.

Beenken K, Dunman P, McAleese F, Macapagal D, Murphy E, Projan S, Blevins J, Smeltzer M. Global gene expression in *Staphylococcus aureus* biofilms. Journal of Bacteriology. 2004; 186(14): 4665-4684.

Begg K, Dewar S, Donachie W. A new *Escherichia coli* cell division gene, *ftsk*. Journal of Bacteriology. 1995; 177(21): 6211-6222.

Berglund C, Ito T, Ikeda M, Ma X, Soderquist B, Hiramatsu K. Novel type of staphylococcal cassette chromosome *mec* in a methicillin-resistant *Staphylococcus aureus* strain isolated in Sweden. Antimicrobial Agents and Chemotherapy. 2008; 52(10): 3512-3516.

Berglund C, Ito T, Ma X, Ikeda M, Watanabe S, Söderquist B, Hiramatsu K. Genetic diversity of methicillin-resistant *Staphylococcus aureus* carrying type IV SCC*mec* in Örebro county and the western region of Sweden. Journal of Antimicrobial Chemotherapy. 2009; 63(1): 32-41.

Besier S, Ludwig A, Brade V, Wichelhaus T. Molecular analysis of fusidic acid resistance in *Staphylococcus aureus*. Molecular Microbiology. 2003; 47(2): 463-469.

Betley M, Mekalanos J. Staphylococcal enterotoxin A is encoded by phage. Science. 1985; 229(4709): 185-187.

Bhakdi S, Tranum-Jensen J. Alpha-toxin of *Staphylococcus aureus*. Microbiological Reviews. 1991; 55(4): 733-751.

Bikandi J, San Millán R, Rementeria A, Garaizar J. *In silico* analysis of complete bacterial genomes: PCR, AFLP-PCR and endonuclease restriction. Bioinformatics. 2004; 20(5): 798-799.

Bingen E, Denamur E, Elion J. Use of ribotyping in epidemiological surveillance of nosocomial outbreaks. Clinical Microbiology Reviews. 1994; 7(3): 311-327.

Bittar F, Ouchenane Z, Smati F, Raoult D, Rolain J. MALDI-TOF-MS for rapid detection of staphylococcal Panton-Valentine leukocidin. International Journal of Antimicrobial Agents. 2009; 34(5): 467-470.

Blair J, Williams R. Phage typing of staphylococci. Bulletin of the World Health Organisation. 1961; 24(6): 771-784.

Boakes E, Kearns A, Ganner M, Perry C, Warner M, Hill R, Ellington M. Molecular diversity within clonal complex 22 methicillin-resistant *Staphylococcus aureus* encoding Panton-Valentine leucocidin in England and Wales. Clinical Microbiology and Infection. 2011; 17(2): 140-145.

Bohach G, Foster T. *Staphylococcus aureus* exotoxins. Fischetti V, Novick R, Ferreti J, Portnoy D, Rood J, editors. Washington D. C: American Society for Microbiology; 2000.

Borrell S, Thorne N, Español M, Mortimer C, Orcau A, Coll P, Gharbia S, González-Martín J, Arnold C. Comparison of four-colour IS*6110*-fAFLP with the classic IS*6110*-RFLP on the ability to detect recent transmission to the city of Barcelona, Spain. Tuberculosis. 2009; 89(3): 233-237.

Boundy S, Safo M, Wang L, Musayev F, O'Farrell H, Rife J, Archer G. Characterization of the *Staphylococcus aureus* rRNA methyltransferase encoded by *orfX*, the gene containing the staphylococcal chromosome cassette *mec* (SCC*mec*) insertion site. The Journal of Biological Chemistry. 2013; 288(1): 132-140.

British Society for Antimicrobial Chemotherapy. BSAC Bacteraemia Resistance Surveillance Programme. [Online]; 2010 [cited 2010]. Available from: www.bsacsurv.org.

Buck J, Como-sabetti K, Harriman K, Danila R, Boxrud D, Glennen A, Lynfield R. Community-associated methicillin-resistant *Staphylococcus aureus*. Emerging Infectious Diseases. 2005; 11(10): 2000-2003.

Bugg T, Wright G, Dutka-Malen S, Arther M, Courvalin P, Walsh C. Molecular basis of vancomycin resistance in *Enterococcus faecium* BM4147: biosynthesis of a depsipeptide peptidoglycan precursor by vancomycin resistance proteins VanH and VanA. Biochemistry. 1991; 30(43): 10408-10415.

Busch U, Nitschko H. Methods for the differentiation of microorganisms. Journal of Chromatography. Biomedical Sciences and Application. 1999; 722(1-2): 263-278.

Butler M, Skow D, Stephenson R, Lyden P, LaMarr W, Foster K. Low frequencies of resistance among *Staphylococcus* and *Enterococcus* species to the bactericidal DNA polymerase inhibitor N(3)-hydroxybutyl 6-(3'-ethyl-4'-methylanilino) uracil. Antimicrobial Agents and Chemotherapy. 2002; 46(12): 3770-3775.

Cai L, Kong F, Wang Q, Wang H, Xiao M, Sintchenko V, Gilbert G. A new multiplex PCR-based reverse line-blot hybridization (mPCR/RLB) assay for rapid staphylococcal cassette chromosome *mec* (SCC*mec*) typing. Journal of Medical Microbiology. 2009; 58(Part 8): 1045-1057.

Cameron A, Eggers D. An ion "Velocitron". The Review of Scientific Instruments. 1948; 19(9): 605-607.

Campo N, Dias M, Daveran-Mingot M, Ritzenthaler P, Le Bourgeois P. Chromosomal constraints in Gram-positive bacteria revealed by artificial inversions. Molecular Microbiology. 2004; 51(2): 511-522.

Cantón R. Antibiotic resistance genes from the environment: a perspective through newly identified antibiotic resistance mechanisms in the clinical setting. Clinical Microbiology and Infection. 2009; 15(Supplement 1): 20-25.

Carbonnelle E, Beretti J, Cottyn S, Quesne G, Berche P, Nassif X, Ferroni A. Rapid identification of staphylococci isolated in clinical microbiology laboratories by matrix-assisted laser desorption ionization-time of flight mass spectrometry. Journal of Clinical Microbiology. 2007; 45(7): 2156-2161.

Carles-Nurit M, Christophle B, Broche S, Gouby A, Bouziges N, Ramuz M. DNA polymorphisms in methicillin-susceptible and methicillin-resistant strains of *Staphylococcus aureus*. Journal of Clinical Microbiology. 1992; 30(8): 2092-2096.

Carroll D, Kehoe M, Cavanagh D, Coleman D. Novel organization of the site-specific integration and excision recombination functions of the *Staphylococcus aureus* serotype F virulence-converting phages phi 13 and phi 42. Molecular Microbiology. 1995; 16(5): 877-893.

Carroll R, Robison T, Rivera F, Davenport J, Jonsson I, Florczyk D, Tarkowski A, Potempa J, Koziel J, Shaw L. Identification of an intracellular M17 family leucine aminopeptidase that is required for virulence in *Staphylococcus aureus*. Microbes and Infection. 2012 ; 14(11): 989-999.

Cassat J, Dunman P, Murphy E, Projan S, Beenken K, Palm K, Yang S, Rice K, Bayles K, Smeltzer M. Transcriptional profiling of a *Staphylococcus aureus* clinical isolate and its isogenic *agr* and *sarA* mutants reveals global differences in comparison to the laboratory strain RN6390. Microbiology. 2006; 152(Part 10): 3075-3090.

Chambers H. The changing epidemiology of *Staphylococcus aureus*. Emerging Infectious Diseases. 2001; 7(2): 178-182.

Chavakis T, Hussain M, Kanse S, Peters G, Bretzel R, Flock J, Herrmann M, Preissner K. *Staphylococcus aureus* extracellular adherence protein serves as anti-inflammatory factor by inhibiting the recruitment of host leukocytes. Nature Medicine. 2002; 8(7): 687-693.

Chen L, Mediavilla J, Oliveira D, Willey B, de Lencastre H, Kreiswirth B. Multiplex real-time PCR for rapid staphylococcal cassette chromosome *mec* typing. Journal of Clinical Microbiology. 2009; 47(11): 3692-3706.

Cheng G, Shapir N, Sadowsky M, Wackett L. Allophanate hydrolase, not urease, functions in bacterial cyanuric acid metabolism. Applied and Environmental Microbiology. 2005; 71(8): 4437-4445.

Chiu Y, Lo W, Wang C. Risk factors and molecular analysis of Panton-Valentine leukocidin-positive methicillin-susceptible *Staphylococcus aureus* colonization and infection in children. Journal of Microbiology, Immunology, and Infection. 2011; 45(3): 208-213.

Chua K, Seemann T, Harrison P, Davies J, Coutts S, Chen H, Haring V, Moore R, Howden B, Stinear T. Complete genome sequence of *Staphylococcus aureus* strain JKD6159, a unique Australian clone of ST93-IV community methicillin-resistant *Staphylococcus aureus*. Journal of Bacteriology. 2010; 192(20): 5556-5557.

Chua T, Moore C, Perri M, Donabedian S, Masch W, Vager D, Davis S, Lulek K, Zimnicki B, Zervos M. Molecular epidemiology of methicillin-resistant *Staphylococcus aureus* bloodstream isolates in urban Detroit. Journal of Clinical Microbiology. 2008; 46(7): 2345-2352.

Clopper C, Pearson E. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika. 1934; 26(4): 404-413.

Cockfield J, Pathak S, Edgeworth J, Lindsay J. Rapid determination of hospital-acquired meticillin- resistant *Staphylococcus aureus* lineages. Journal of Medical Microbiology. 2007; 56(Part 5): 614-619.

Coenye T, Falsen E, Vancanneyt M, Hoste B, Govan J, Kersters K, Vandamme P. Classification of *Alcaligenes faecalis*-like isolates from the environment and human clinical samples as *Ralstonia gilardii* sp. *nov*. International Journal of Systematic and Evolutionary Microbiology. 1999; 49(2): 405-413.

Conceição T, de Sousa M, de Lencastre H. Staphylococcal interspersed repeat unit typing of *Staphylococcus aureus*: evaluation of a new multilocus variable-number tandem-repeat analysis typing method. Journal of Clinical Microbiology. 2009; 47(5): 1300-1308.

Cooke F, Brown N. Community-associated methicillin-resistant *Staphylococcus aureus* infections. British Medical Bulletin. 2010; 94(1): 215-227.

Cookson B. Five decades of MRSA: controversy and uncertainty continues. The Lancet. 2011; 378(9799): 1291-1292.

Cosgrove S, Sakoulas G, Perencevich E, Schwaber M, Karchmer A, Carmeli Y. Comparison of mortality associated with methicillin-resistant and methicillin-susceptible *Staphylococcus aureus* bacteremia: a meta-analysis. Clinical Infectious Diseases. 2003; 36(1): 53-59.

Costerton J, Stewart P, Greenberg E. Bacterial biofilms: a common cause of persistent infections. Science. 1999; 284(5418): 1318-1322.

Couto I, Sanches I, Sá-Leão R, de Lencastre H. Molecular characterization of *Staphylococcus sciuri* strains isolated from humans. Journal of Clinical Microbiology. 2000; 38(3): 1136-1143.

Cox R, Conquest C, Mallaghan C, Marples R. A major outbreak of methicllin-resistant *Staphylococcus aureus* caused by a new phage-type (EMRSA-16). Journal of Hospital Infection. 1995; 29: 87-106.

Crisóstomo M, Westh H, Tomasz A, Chung M, Oliveira D, de Lencastre H. The evolution of methicillin resistance in *Staphylococcus aureus*: similarity of genetic backgrounds in historically early methicillin-susceptible and -resistant isolates and contemporary epidemic clones. Proceedings for the National Academy of Science. 2001; 98(17): 9865-9870.

Cucarella C, Solano C, Valle J, Amorena B, Lasa I, Penadés J. Bap, a *Staphylococcus aureus* surface protein involved in biofilm formation. Journal of Bacteriology. 2001; 183(9): 2888-2896.

Cui L, Murakami H, Kuwahara-Arai K, Hanaki H, Hiramatsu K. Contribution of a thickened cell wall and its glutamine nonamidated component to the vancomycin resistance expressed by *Staphylococcus aureus* Mu50. Antimicrobial Agents and Chemotherapy. 2000; 44(9): 2276-2285.

Dale J, Park S. Molecular genetics of bacteria Chichester: John Wiley & Sons Ltd; 2004.

Damle P, Wall E, Spilman M, Dearborn A, Ram G, Novick R, Dokland T, Christie G. The roles of SaPI proteins gp7 (CpmA) and gp6 (CpmB) in capsid size determination and helper phage interference. Virology. 2012; 432(2): 277-282.

Datta N, Hughes V. Plasmids of the same Inc groups in Enterobacteria before and after the medical use of antibiotics. Nature. 1983; 306(5943): 616-617.

David M, Daum R. Community-associated methicillin-resistant *Staphylococccus aureus*: epidemiology and clinical consequences of an emerging epidemic. Clinical Microbiology Reviews. 2010; 23(3): 616-687.

D'Costa V, King C, Kalan L, Morar M, Sung W, Schwarz C, Froese D, Zazula G, Calmels F, Debruyne R, Golding G, Poinar H, Wright G. Antibiotic resistance is ancient. Nature. 2011; 477(7365): 457-461.

de Haas C, Veldkamp K, Peschel A, Weerkamp F, van Wamel W, Heezius E, Poppelier M, van Kessel K, van Strijp J. Chemotaxis inhibitory protein of *Staphylococcus aureus*, a bacterial antiinflammatory agent. The Journal of Experimental Medicine. 2004; 199(5): 687-695.

de Lencastre H, Chung M, Westh H. Archaic strains of methicillin-resistant *Staphylococcus aureus*: molecular and microbiological properties of isolates from the 1960s in Denmark. Microbial Drug Resistance. 2000; 6(1): 1-10.

de Lencastre H, Tomasz A. Reassessment of the number of auxiliary genes essential for expression of high-level methicillin resistance in *Staphylococcus aureus*. Antimicrobial Agents and Chemotherapy. 1994; 38(11): 2590-2598.

de Lencastre H, Wu S, Pinho M, Ludovice A, Filipe S, Gardete S, Sobral R, Gill S, Chung M, Tomasz A. Antibiotic resistance as a stress response: complete sequencing of a large number of chromosomal loci in *Staphylococcus aureus* strain COL that impact on the expression of resistance to methicillin. Microbial Drug Resistance. 1999; 5(3): 163-175.

de Vos P, Garrity G, Jones D, Krieg N, Ludwig W, Rainey F, Schleifer K.-H, Whitman W. editors. Genus I. *Staphylococcus*. In Bergey's manual of systematic bacteriology. 2nd ed. New York: Springer; 2009.

Dempsey R, Carroll D, Kong H, Higgins L, Keane C, Coleman D. *Sau*42I, a *Bcg*I-like restriction-modification system encoded by the *Staphylococcus aureus* quadruple-converting phage φ42. Microbiology. 2005; 151(Part 4): 1301-1311.

Desai M, Efstratiou A, George R, Stanley J. High-resolution genotyping of *Streptococcus pyogenes* serotype M1 isolates by fluorescent amplified-fragment length polymorphism analysis. Journal of Clinical Microbiology. 1999; 37(6): 1948-1952.

Desai M, Logan J, Frost J, Stanley J. Genome sequence-based fluorescent amplified fragment length polymorphism of *Campylobacter jejuni*, its relationship to serotyping, and its implications for epidemiological analysis. Journal of Clinical Microbiology. 2001; 39(11): 3823-3829.

Desai M, Tanna A, Wall R, Efstratiou A, George R, Stanley J. Fluorescent amplified-fragment length polymorphism analysis of an outbreak of group A streptococcal invasive disease. Journal of Clinical Microbiology. 1998; 36(11): 3133-3137.

Desai M. Molecular epidemiological typing of *Streptococcus pyogenes*. PhD thesis. Open University. 1999.

Diep B, Gill S, Chang R, Phan T, Chen J, Davidson M, Lin F, Lin J, Carleton H, Mongodin E, Sensabaugh G, Perdreau-Remington F. Complete genome sequence of USA300, an epidemic clone of community-acquired meticillin-resistant *Staphylococcus aureus*. The Lancet. 2006; 367(9512): 731-739.

Diep B, Otto M. The role of virulence determinants in community-associated MRSA pathogenesis. Trends in Microbiology. 2008; 16(8): 361-369.

Diep B, Stone G, Basuino L, Graber C, Miller A, des Etages S, Jones A, Palazzolo-Ballance A, Perdreau-Remington F, Sensabaugh G, DeLeo F, Chambers H. The arginine catabolic mobile element and staphylococcal chromosomal cassette *mec* linkage: convergence of virulence and resistance in the USA300 clone of methicillin-resistant *Staphylococcus aureus*. The Journal of Infectious Diseases. 2008; 197(11): 1523-1530.

Dintner S, Staroń A, Berchtold E, Petri T, Mascher T, Gebhard S. Coevolution of ABC transporters and two-component regulatory systems as resistance modules against antimicrobial peptides in Firmicutes bacteria. Journal of Bacteriology. 2011; 193(15): 3851-3862.

Duckworth D, Glenn J, McCorguodale D. Inhibition of bacteriophage replication by extrachromosomal genetic elements. Microbiological Reviews. 1981; 45(1): 52-71.

Edwards U, Rogall T, Blöcker H, Emde M, Böttger E. Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. Nucleic Acids Research. 1989; 17(19): 7843-7853.

Ellington M, Hope R, Livermore D, Kearns A, Henderson K, Cookson B, Pearson A, Johnson A. Decline of EMRSA-16 amongst methicillin-resistant *Staphylococcus aureus* causing bacteraemias in the UK between 2001 and 2007. Journal of Antimicrobial Chemotherapy. 2010; 65(3): 446-448.

Ellington M, Yearwood L, Ganner M, East C, Kearns A. Distribution of the ACME-*arcA* gene among methicillin-resistant *Staphylococcus aureus* from England and Wales. Journal of Antimicrobial Chemotherapy. 2008; 61(1): 73-77.

Emonts M, Uitterlinden A, Nouwen J, Kardys I, de Maat M, Melles D, Witteman J, de Jong P, Verbrugh H, Hofman A, Hermans P, van Belkum A. Host polymorphisms in interleukin 4, complement factor H, and C-reactive protein associated with nasal carriage of *Staphylococcus aureus* and occurrence of boils. The Journal of Infectious Diseases. 2008; 197(9): 1244-1253.

Engelberg-Kulka H, Glaser G. Addiction modules and programmed cell death and antideath in bacterial cultures. Annual Review of Microbiology. 1999; 53: 43-70.

England R, Pettersson M. Pyro Q-CpGTM: quantitative analysis of methylation in multiple CpG sites by pyrosequencing®. Nature Methods. 2005; 2(10).

Enright M, Day N, Davies C, Peacock S, Spratt B. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. Journal of Clinical Microbiology. 2000; 38(3): 1008-1015.

Enright M, Robinson D, Randle G, Feil E, Grundmann H, Spratt B. The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA). Proceedings for the National Academy of Sciences. 2002; 99(11): 7687-7692.

Enright M. The evolution of a resistant pathogen-the case of MRSA. Current Opinion in Pharmacology. 2003; 3(5): 474-479.

Erni B, Siebold C, Christen S, Srinivas A, Oberholzer A, Baumann U. Small substrate, big surprise: fold, function and phylogeny of dihydroxyacetone kinases. Cellular and Molecular Life Sciences. 2006; 63(7-8): 890-900.

Evans J, Bradford W, Niven C. Comments concerning the taxonomy of the genera *Micrococcus* and *Staphylococcus*. International Bulletin of Bacteriological Nomenclature and Taxonomy. 1955; 5(2): 61-66.

Fakhrai-Rad H, Pourmand N, Ronaghi M. Pyrosequencing™: an accurate detection platform for single nucleotide polymorphisms. Human Mutation. 2002; 19(5): 476-485.

Feil E, Cooper J, Grundmann H, Robinson D, Enright M, Berendt T, Peacock S, Smith J, Murphy M, Spratt B, Moore C, Day N. How clonal is *Staphylococcus aureus*? Journal of Bacteriology. 2003; 185(11): 3307-3316.

Feil E, Enright M. Analyses of clonality and the evolution of bacterial pathogens. Current Opinion in Microbiology. 2004; 7(3): 308-313.

Feil E, Li B, Aanensen D, Hanage W, Spratt B. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. Journal of Bacteriology. 2004; 186(5): 1518-1530.

Fitzgerald J, Sturdevant D, Mackie S, Gill S, Musser J. Evolutionary genomics of *Staphylococcus aureus*: insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. Proceedings for the National Academy of Sciences. 2001; 98(15): 8821-8826.

Fitzpatrick M, Burggrave R. Optical Mapping: A novel single molecule system for microbial comparative genomics and genome sequence assembly. 2009.

Flannagan S, Chow J, Donabedian S, Brown W, Perri M, Zervos M, Ozawa Y, Clewell D. Plasmid content of a vancomycin-resistant *Enterococcus faecalis* isolate from a patient also colonized by *Staphylococcus aureus* with a VanA phenotype. Antimicrobial Agents and Chemotherapy. 2003; 47(12): 3954-3959.

Fluit A, Schmitz F. MRSA current perspectives. Norfolk: Caister Academic Press; 2003.

Fode P, Larsen A, Feenstra B, Jespersgaard C, Skov R, Stegger M, Fowler V, Danish SAB Study Group Consortium, Andersen P. Genetic variability in beta-defensins is not associated with susceptibility to *Staphylococcus aureus* bacteremia. Public Library of Science One. 2012; 7(2): e32315.

Fowler V, Sakoulas G, McIntyre L, Meka V, Arbeit R, Cabell C, Stryjewski M, Eliopoulos G, Reller L, Corey G, Jones T, Lucindo N, Yeaman M, Bayer A. Persistent bacteremia due to methicillin-resistant *Staphylococcus aureus* infection is associated with *agr* dysfunction and low-level in vitro resistance to thrombin-induced platelet microbicidal protein. The Journal of Infectious Diseases. 2004; 190(6): 1140-1149.

Frank A, Marcinak J, Mangat P, Schreckenberger P. Increase in Community-acquired methicillin-resistant *Staphylococcus aureus* in children. Clinical Infectious Diseases. 1999; 29(4): 935-936.

Frankel M, Wojcik B, DeDent A, Missiakas D, Schneewind O. ABI domain-containing proteins contribute to surface protein display and cell division in *Staphylococcus aureus*. Molecular Microbiology. 2010; 78(1): 238-252.

Frère J. Beta-lactamases and bacterial resistance to antibiotics. Molecular Microbiology. 1995; 16(3): 385-395.

Fujimura T, Murakami K. *Staphylococcus aureus* clinical isolate with high-level methicillin resistance with an *lytH* mutation caused by IS*1182* insertion. Antimicrobial Agents and Chemotherapy. 2008; 52(2): 643-647.

Gagnaire J, Dauwalder O, Boisset S, Khau D, Freydière A, Ader F, Bes M, Lina G, Tristan A, Reverdy M, and A, Geissmann T, Benito Y, Durand G, Charrier J, Etienne J, Welker M, van Belkum A, Vandenesch F. Detection of *Staphylococcus aureus* delta-toxin production by whole-cell MALDI-TOF mass spectrometry. Public Library of Science One. 2012; 7(7): e40660.

Ganga R, Riederer K, Sharma M, Fakih M, Johnson L, Shemes S, Khatib R. Role of SCC*mec* type in outcome of *Staphylococcus aureus* bacteremia in a single medical center. Journal of Clinical Microbiology. 2009; 47(3): 590-595.

García-Álvarez L, Holden M, Lindsay H, Webb C, Brown D, Curran M, Walpole E, Brooks K, Pickard D, Teale C, Parkhill J, Bentley S, Edwards G, Girvan E, Kearns A, Pichon B, Hill R, Larsen A, Skov R, Peacock S, Maskell D, Holmes M. Meticillin-resistant *Staphylococcus aureus* with a novel *mecA* homologue in human and bovine populations in the UK and Denmark: a descriptive study. The Lancet Infectious Diseases. 2011; 11(8): 595-603.

Garvey P, van Sinderen D, Twomey D, Hill C, Fitzgerald G. Molecular genetics of bacteriophage and natural phage defence systems in the genus *Lactococcus*. International Dairy Journal. 1995; 5(8): 905-947.

Ge M, Chen Z, Onishi H, Kohler J, Silver L, Kerns R, Fukuzawa S, Thompson C, Kahne D. Vancomycin derivatives that inhibit peptidoglycan biosynthesis without binding D-Ala-D-Ala. Science. 1999; 284(5413): 507-511.

Geissler B, Margolin W. Evidence for functional overlap among multiple bacterial cell division proteins: compensating for the loss of FtsK. Molecular Microbiology. 2005; 58(2): 596-612.

Gharizadeh B, Nordström T, Ahmadian A, Ronaghi M, Nyrén P. Long-read pyrosequencing using pure 2'-deoxyadenosine-5'-O'-(1-thiotriphosphate) Sp-isomer. Analytical Biochemistry. 2002; 301(1): 82-90.

Ghuysen J. Serine β-lactamases and penicillin-binding proteins. Annual Review of Microbiology. 1991; 45: 37-67.

Gill S, Fouts D, Archer G, Mongodin E, DeBoy R, Ravel J, Paulsen I, Kolonay J, Brinkac L, Beanan M, Dodson R, Daugherty S, Madupu R, Angiuoli S, Durkin A, Haft D, Vamathevan J, Khouri H, Utterback T, Lee C, Dimitrov G, Jiang L, Qin H, Weidman J, Tran K, Kang K, Hance I, Nelson K, Fraser C. Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant *Staphylococcus aureus* strain and a biofilm-producing methicillin-resistant *Staphylococcus* epidermidis strain. Journal of Bacteriology. 2005; 187(7): 2426-2438.

Gillet Y, Issartel B, Vanhems P, Fournet J, Lina G, Bes M, Vandenesch F, Piémont Y, Brousse N, Floret D, Etienne J. Association between *Staphylococcus aureus* strains carrying gene for Panton-Valentine leukocidin and highly lethal necrotising pneumonia in young immunocompetent patients. The Lancet. 2002; 359(9308): 753-759.

Goering R, Morrison D, Al-Doori Z, Edwards G, Gemmell C. Usefulness of *mec*-associated direct repeat unit (*dru*) typing in the epidemiological analysis of highly clonal methicillin-resistant *Staphylococcus aureus* in Scotland. Clinical Microbiology and Infection. 2008; 14(10): 964-969.

Goerke C, Kümmel M, Dietz K, Wolz C. Evaluation of intraspecies interference due to *agr* polymorphism in *Staphylococcus aureus* during infection and colonization. The Journal of Infectious Diseases. 2003; 188(2): 250-256.

Gomes A, Westh H, de Lencastre H. Origins and evolution of methicillin-resistant *Staphylococcus aureus* clonal lineages. Antimicrobial Agents and Chemotherapy. 2006; 50(10): 3237-3244.

Gonzales T, Robert-Baudouy J. Bacterial aminopeptidases: properties and functions. Federation of European Microbiological Societies Microbiology Reviews. 1996; 18(4): 319-344.

Goodyear C, Silverman G. Staphylococcal toxin induced preferential and prolonged *in vivo* deletion of innate-like B lymphocytes. Proceedings for the National Academy of Sciences. 2004; 101(31): 11392-11397.

Gort G, van Eeuwijk F. Review and simulation of homoplasy. Euphytica. 2012; 183(3): 389-400.

Gottesman S. Micros for microbes: non-coding regulatory RNAs in bacteria. Trends in Genetics. 2005; 21(7): 399-404.

Gould S, Rollason J, Hilton A, Cuschieri P, McAuliffe L, Easmon S, Fielder M. UK epidemic strains of meticillin-resistant *Staphylococcus aureus* in clinical samples from Malta. Journal of Medical Microbiology. 2008; 57(Part 11): 1394-1398.

Goulding J, Hookey J, Stanley J, Olver W, Neal K, Ala'Aldeen D, Arnold C. Fluorescent amplified-fragment length polymorphism genotyping of *Neisseria meningitidis* identifies clones associated with invasive disease. Journal of Clinical Microbiology. 2000; 38(12): 4580-4585.

Grady R, Desai M, O'Neill G, Cookson B, Stanley J. Genotyping of epidemic methicillin-resistant *Staphylococcus aureus* phage type 15 isolates by fluorescent amplified- fragment length polymorphism analysis. Journal of Clinical Microbiology. 1999; 37(10): 3198-3203.

Grady R, Hayes F. Axe-Txe, a broad-spectrum proteic toxin-anttoxin system specified by a mutlidrug-resistant, clinical isolate of *Enterococcus faecium*. Molecular Microbiology. 2003; 47(5): 1419-1432.

Grady R, O'Neill G, Cookson B, Stanley J. Fluorescent amplified-fragment length polymorphism analysis of the MRSA epidemic. Federation of European Microbiological Societies Microbiology Letters. 2000; 187(1): 27-30.

Gravat A, Colin D, Keller D, Girardot R, Monteil H, Prévost G, Giradot R. Characterization of a novel structural member, LukE-LukD, of the bi-component staphylococcal leucotoxins family. FEBS letters. 1998; 436(2): 202-208.

Griffin T, Parsons L, Leschziner A, DeVost J, Derbyshire K, Grindley N. *In vitro* transposition of Tn*552*: a tool for DNA sequencing and mutagenesis. Nucleic Acids Research. 1999; 27(19): 3859-3865.

Groicher K, Firek B, Fujimoto D, Bayles K. The *Staphylococcus aureus lrgAB* operon modulates murein hydrolase activity and penicillin tolerance. Journal of Bacteriology. 2000; 182(7): 1794-1801.

Grundmann H, Aanensen D, van den Wijngaard C, Spratt B, Harmsen D, Friedrich A. Geographic distribution of *Staphylococcus aureus* causing invasive infections in Europe: a molecular-epidemiological analysis. Public Library of Science Medicine. 2010; 7(1): e1000215.

Grundmann H, Hori S, Tanner G. Determining confidence intervals when measuring genetic diversity and the discriminatory abilities of typing methods for microorganisms. Journal of Clinical Microbiology. 2001; 39(11): 4190-4192.

Grundmeier M, Hussain M, Becker P, Heilmann C, Peters G, Sinha B. Truncation of fibronectin-binding proteins in *Staphylococcus aureus* strain Newman leads to deficient adherence and host cell invasion due to loss of the cell wall anchor function. Infection and Immunity. 2004; 72(12): 7155-7163.

Hanberger H, Walther S, Leone M, Barie P, Rello J, Lipman J, Marshall J, Anzueto A, Sakr Y, Pickkers P, Felleiter P, Engoren M, Vincent J. Increased mortality associated with methicillin-resistant *Staphylococcus aureus* (MRSA) infection in the intensive care unit: results from the EPIC II study. International Journal of Antimicrobial Agents. 2011; 38(4): 331-335.

Hanssen A, Sollid J. SCC*mec* in staphylococci: genes on the move. Federation of European Microbiological Societies Immunology and Medical Microbiology. 2006; 46(1): 8-20.

Hardy K, Oppenheim B, Gossain S, Gao F, Hawkey P. Use of the variations in staphylococcal interspersed repeat units for molecular typing of methicillin-resistant *Staphylococcus aureus* strains. Journal of Clinical Microbiology. 2006; 44(1): 271-273.

Hardy K, Ussery D, Oppenheim B, Hawkey P. Distribution and characterization of staphylococcal interspersed repeat units (SIRUs) and potential use for strain differentiation. Microbiology. 2004; 150(Part 12): 4045-4052.

Harraghy N, Hussain M, Haggar A, Chavakis T, Sinha B, Herrmann M, Flock J. The adhesive and immunomodulating properties of the multifunctional *Staphylococcus aureus* protein Eap. Microbiology. 2003; 149(10): 2701-2707.

Harris L, Foster S, Richards R. An introduction to *Staphylococcus aureus*, and techniques for identifying and quantifying *S. aureus* adhesins in relation to adhesion to biomaterials: Review. European Cells and Materials. 2002; 4: 39-60.

Harris S, Cartwright E, Török M, Holden M, Brown N, Ogilvy-Stuart A, Ellington M, Quail M, Bentley S, Parkhill J, Peacock S. Whole-genome sequencing for analysis of an outbreak of methicillin-resistant *Staphylococcus aureus*: a descriptive study. The Lancet Infectious Diseases. 2013; 13(2): 130-136.

Hartman B, Tomasz A. Altered penicillin-binding proteins in methicillin-resistant strains of *Staphylococcus aureus*. Antimicrobial Agents and Chemotherapy. 1981; 19(5): 726-735.

Hartman B, Tomasz A. Low-affinity penicillin-binding protein associated with β-lactam resistance in *Staphylococcus aureus*. Journal of Bacteriology. 1984; 158(2): 513-516.

Hartstein A, Phelps C, Kwok R, Mulligan M. *In vivo* stability and discriminatory power of methicillin-resistant *Staphylococcus aureus* typing by restriction endonuclease analysis of plasmid DNA compared with those of other molecular methods. Journal of Clinical Microbiology. 1995; 33(8): 2022-2026.

Hawkey P. The origins and molecular basis of antibiotic resistance. British Medical Journal. 1998; 317(7159): 657-660.

Henderson I, Owen P, Nataro J. Molecular switches - the ON and OFF of bacterial phase variation. Molecular Microbiology. 1999; 33(5): 919-932.

Hershberg R, Petrov D. Selection on codon bias. Annual Review of Genetics. 2008; 42: 287-299.

Heusser R, Ender M, Berger-Bächi B, McCallum N. Mosaic staphylococcal cassette chromosome *mec* containing two recombinase loci and a new *mec* complex, B2. Antimicrobial Agents and Chemotherapy. 2007; 51(1): 390-393.

Hibbing M, Fugua C, Parsek M, Peterson S. Bacterial competition: surviving and thriving in the microbial jungle. Nature Reviews Microbiology. 2010; 8(1): 15-25.

Highlander S, Hultén K, Qin X, Jiang H, Yerrapragada S, Mason E, Shang Y, Williams T, Fortunov R, Liu Y, Igboeli O, Petrosino J, Tirumalai M, Uzman A, Fox G, Cardenas A, Muzny D, Hemphill L, Ding Y, Dugan S, Blyth P, Buhay C, Dinh H, Hawes A, Holder M, Kovar C, Lee S, Liu W, Nazareth L, Wang Q, Zhou J, Kaplan S, Weinstock G. Subtle genetic changes enhance virulence of methicillin resistant and sensitive *Staphylococcus aureus*. BioMed Central Microbiology. 2007; 7: 99.

Higuchi W, Takano T, Teng L, Yamamoto T. Structure and specific detection of staphylococcal cassette chromosome *mec* type VII. Biochemical and Biophysical Research Communications. 2008; 377(3): 752-756.

Hill L. The adansonian classification of the staphylococci. Journal of General Microbiology. 1959; 20(2): 277-283.

Hiramatsu K, Cui L, Kuroda M, Ito T. The emergence and evolution of methicillin-resistant *Staphylococcus aureus*. Trends in Microbiology. 2001; 9(10): 486-493.

Hiramatsu K, Hanaki H, Ino T, Yabuta K, Oguri T, Tenover F. Methicillin-resistant *Staphylococcus aureus* clinical strain with reduced vancomycin susceptibility. Journal of Antimicrobial Chemotherapy. 1997; 40(9): 135-146.

Hofer M, Newell K, Duke R, Schlievert P, Freed J, Leung D. Differential effects of staphylococcal toxic shock syndrome toxin-1 on B cell apoptosis. Proceedings for the National Academy of Sciences. 1996; 93(11): 5425-5430.

Holden M, Feil E, Lindsay J, Peacock S, Day N, Enright M, Foster T, Moore C, Hurst L, Atkin R, Barron A, Bason N, Bentley S, Chillingworth C, Chillingworth T, Churcher C, Clark L, Corton C, Cronin A, Doggett J, Dowd L, Feltwell T, Hance Z, Harris B, Hauser H, Holroyd S, Jagels K, James K, Lennard N, Line A, es R, Moule S, Mungall K, Ormond D, Quail, M, Rabbinowitsch E, Rutherford K, Sanders M, Sharp S, Simmonds M, Stevens K, Whitehead S, Barrell B, Spratt B, Parkill J. Complete genomes of two clinical *Staphylococcus aureus* strains: evidence for the rapid evolution of virulence and drug resistance. Proceedings for the National Academy of Sciences. 2004; 101(26): 9786-9791.

Holden M, Lindsay J, Corton C, Quail M, Cockfield J, Pathak S, Batra R, Parkhill J, Bentley S, Edgeworth J. Genome sequence of a recently emerged, highly transmissible, multi-antibiotic- and antiseptic-resistant variant of methicillin-resistant *Staphylococcus aureus*. Journal of Bacteriology. 2010; 192(3): 888-892.

Holden M, Lindsay J, Enright M, Francois P, Schrenzel J, McNamara P, Fournier B, Horsburgh M, Pinho M, Queck S, Otto M, Fitzgerald J, Penadés J. *Staphylococcus* molecular genetics. Linday J, editor. Norfolk: Caister academic press; 2008.

Holmes A, Edwards G, Girvan E, Hannant W, Danial J, Fitzgerald J, Templeton K. Comparison of two multilocus variable-number tandem-repeat methods and pulsed-field gel electrophoresis for differentiating highly clonal methicillin-resistant *Staphylococcus aureus* isolates. Journal of Clinical Microbiology. 2010; 48(10): 3600-3607.

Honda M, Fujisawa T, Shibata T, Mikawa T. RecR forms a ring-like tetramer that encircles dsDNA by forming a complex with RecF. Nucleic Acids Research. 2008; 36(15): 5013-5020.

Hookey J, Edwards V, Patel S, Richardson J, Cookson B. Use of fluorescent amplified fragment length polymorphism (fAFLP) to characterise methicillin-resistant *Staphylococcus aureus*. Journal of Microbiological Methods. 1999; 37(1): 7-15.

Hope R, Livermore D, Brick G, Lillie M, Reynolds R. Non-susceptibilities trends among staphylococci from bacteraemias in the UK and Ireland, 2001-2006. Journal of Antimicrobial Chemotherapy. 2008; 62(Supplement 2): 65-74.

Hospital Infection Society and British Society for Antimicrobial Chemotherapy. Revised guidelines for the control of epidemic methicillin-resistant *Staphylococcus aureus*. Report of a combined working party of the Hospital Infection Society and British Society for Antimicrobial Chemotherapy. Journal of Hospital Infection. 1990; 16(4): 351-377.

Hsu L, Loomba-Chlebicka N, Koh Y, Tan T, Krishnan P, Lin R, Tee N, Fisher D, Koh T. Evolving EMRSA-15 epidemic in Singapore hospitals. Journal of Medical Microbiology. 2007; 56(Part 3): 376-379.

Huang T, Chen F, Miu W, Liao T, Lin A, Huang I, Wu K, Tsai S, Chen Y, Lauderdale T. Complete genome sequence of *Staphylococcus aureus* M013, a *pvl*-positive, ST59-SCC*mec* type V strain isolated in Taiwan. Journal of Bacteriology. 2012; 194(5): 1256-1257.

Hunter P, Gaston M. Numerical index of the discriminatory ability of typing systems: an application of Simpson's Index of Diversity. Journal of Clinical Microbiology. 1988; 26(11): 2465-2466.

Ichiyama S, Ohta M, Shimokata K, Kato N, Takeuchi J. Genomic DNA fingerprinting by pulsed-field gel electrophoresis as an epidemiological marker for study of nosocomial infections caused by methicillin-resistant *Staphylococcus aureus*. Journal of Clinical Microbiology. 1991; 29(12): 2690-2695.

Ince D, Hooper D. Quinolone resistance due to reduced target enzyme expression. Journal of Bacteriology. 2003; 185(23): 6883-6892.

Inoue J, Honda M, Ikawa S, Shibata T, Mikawa T. The process of displacing the single-stranded DNA-binding protein from single-stranded DNA by RecO and RecR proteins. Nucleic Acids Research. 2008; 36(1): 94-109.

Inoue R, Kaito C, Tanabe M, Kamura K, Akimitsu N, Sekimizu K. Genetic identification of two distinct DNA polymerases, DnaE and PolC, that are essential for chromosomal DNA replication in *Staphylococcus aureus*. Molecular Genetics and Genomics. 2001; 266(4): 564-571.

International working group on the classification of staphylococcal cassette chromosome elements. Classification of staphylococcal cassette chromosome *mec* (SCC*mec*): guidelines for reporting novel SCC*mec* elements. Antimicrobial Agents and Chemotherapy. 2009; 53(12): 4961-4967.

Ito T, Katayama Y, Asada K, Mori N, Tsutsumimoto K, Tiensasitorn C, Hiramatsu K. Structural comparison of three types of staphylococcal cassette chromosome *mec* integrated in the chromosome in methicillin-resistant *Staphylococcus aureus*. Antimicrobial Agents and Chemotherapy. 2001; 45(5): 1323-1336.

Ito T, Katayama Y, Hiramatsu K. Cloning and nucleotide sequence determination of the entire *mec*DNA of pre-methicillin-resistant *Staphylococcus aureus* strain N315. Antimicrobial Agents and Chemotherapy. 1999; 43(6): 1449-1458.

Ito T, Ma X, Takeuchi F, Okuma K, Yuzawa H, Hiramatsu K. Novel type V staphylococcal cassette chromosome *mec* driven by a novel cassette chromosome recombinase, *ccr*C. Antimicrobial Agents and Chemotherapy. 2004; 48(7): 2637-2651.

Ito T, Okuma K, Ma X, Yuzawa H, Hiramatsu K. Insights on antibiotic resistance of *Staphylococcus aureus* from its whole genome: genomic island SCC. Drug Resistance Updates. 2003; 6(1): 41-52.

Janda J, Abbott S. 16S rRNA gene sequencing for bacterial identification in the diagnostics laboratory: pluses, perils, and pitfalls. Journal of Clinical Microbiology. 2007; 45(9): 2761-2764.

Jarraud S, Peyrat M, Lim A, Tristan A, Bes M, Mougel C, Etienne J, Vandenesch F, Bonneville M, Lina G. *egc*, a highly prevalent operon of enterotoxin gene, forms a putative nursery of superantigens in *Staphylococcus aureus*. The Journal of Immunology. 2001; 166(1): 669-677.

Jayasinghe L, Bayley H. The leukocidin pore: evidence for an octamer with four LukF subunits and four LukS subunits alternating around a central axis. Protein Science. 2005; 14(10): 2550-2561.

Jeffreys A, Wilson V, Thein S. Hypervariable 'minisatellite' regions in human DNA. Nature. 1985; 314(6006): 67-73.

Jevons M, Rolinson G, Knox R. "Celebenin"-resistant staphylococci. British Medical Journal. 1961; 1(5219): 124-126.

Ji G, Beavis R, Novick R. Cell density control of staphylococcal virulence mediated by an octapeptide pheromone. Proceedings of the National Academy of Sciences of the United States of America. 1995; 92(26): 12055-12059.

Johnson A, Aucken H, Cavendish S, Ganner M, Wale M, Warner M, Livermore D, Cookson B. Dominance of EMRSA-15 and -16 among MRSA causing nosocomial bacteraemia in the UK: analysis of isolates from the European Antimicrobial Resistance Surveillance System (EARSS). Journal of Antimicrobial Chemotherapy. 2001; 48: 141-156.

Johnson A, Davies J, Guy R, Abernethy J, Sheridan E, Pearson A, Duckworth G. Mandatory surveillance of methicillin-resistant *Staphylococcus aureus* (MRSA) bacteraemia in England: the first 10 years. Journal of Antimicrobial Chemotherapy. 2012; 67(4): 802-809.

Johnson A, Pearson A, Duckworth G. Surveillance and epidemiology of MRSA bacteraemia in the UK. Journal of Antimicrobial Chemotherapy. 2005; 56(3): 455-462.

Jönsson K, Signäs C, Müller H, Lindberg M. Two different genes encode fibronectin binding proteins in *Staphylococcus aureus*. European Journal of Biochemistry. 1991; 202(3): 1041-1048.

Joubert O, Viero G, Keller D, Martinez E, Colin D, Monteil H, Mourey L, Serra M, Pévost G. Engineered covalent leucotoxin heterodimers form functional pores: insights into S-F interactions. The Biochemical Journal. 2006; 396(2): 381-389.

Jousselin A, Renzoni A, Andrey D, Monod A, Lew D, Kelley W. The postranslocational chaperone lipoprotein prsA is involved in both glycopeptide and oxacillin resistance in *Staphylococcus aureus*. Antimicrobial Agents and Chemotherapy. 2012 ; 56(7): 3629-3640.

Kallen A, Mu Y, Bulens S, Reingold A, Petit S, Gershman K, Ray S, Harrison L, Lynfield R, Dumyati G, Townes J, Schaffner W, Patel P, Fridkin S. Health care-associated invasive MRSA infections, 2005-2008. The Journal of the American Medical Association. 2010; 304(6): 641-648.

Karakawa W, Sutton A, Schneerson R, Karpas A, Vann W. Capsular antibodies induce type-specific phagocytosis of capsulated *Staphylococcus aureus* by human polymorphonuclear leukocytes. Infection and Immunity. 1988; 56(5): 1090-1095.

Katayama Y, Ito T, Hiramatsu K. A new class of genetic element, Staphylococcus cassette chromosome *mec*, encodes methicillin resistance in *Staphylococcus aureus*. Antimicrobial Agents and Chemotherapy. 2000; 44(6): 1549-1555.

Katayama Y, Robinson D, Enright M, Chambers H. Genetic background affects stability of *mecA* in *Staphylococcus aureus*. Journal of Clinical Microbiology. 2005; 43(5): 2380-2383.

Katayama Y, Takeuchi F, Ito T, Ma X, Ui-Mizutani Y, Kobayashi I, Hiramatsu K. Identification in methicillin-susceptible *Staphylococcus hominis* of an active primordial mobile genetic element for the staphylococcal cassette chromosome *mec* of methicillin-resistant *Staphylococcus aureus*. Journal of Bacteriology. 2003; 185(9): 2711-2722.

Katayama Y, Zhang H, Hong D, Chambers H. Jumping the barrier to β-lactam resistance in *Staphylococcus aureus*. Journal of Bacteriology. 2003; 185(18): 5465-5472.

Kawada-Matsuo M, Yoshida Y, Nakamura N, Komatsuzawa H. Role of two-component systems in the resistance of *Staphylococcus aureus* to antibacterial agents. Virulence. 2011; 2(5): 427-430.

Kerr S, Kerr G, Mackintosh C, Marples R. A survey of methicillin-resistant *Staphylococcus aureus* affecting patients in England and Wales. Journal of Hospital Infection. 1990; 16(1): 35-48.

Klein E, Smith D, Laxminarayan R. Hospitalizations and deaths caused by methicillin-resistant *Staphylococcus aureus*, United States, 1999-2005. Emerging Infectious Diseases. 2007; 13(12): 1840-1846.

Kloos W, Ballard D, George C, Webster J, Hubner R, Ludwig W, Schleifer K, Fiedler F. Schubert K. Delimiting the genus *Staphylococcus* through description of *Macrococcus caseolyticus* gen. nov., comb. nov. and *Macrococcus equipercicus* sp. nov., *Macrococcus bovicus* sp. nov. and *Macrococcus carouselicus* sp. nov. International Journal of Systematic Bacteriology. 1998;48(Part 3): 859-877.

Kluytmans J, van Belkum A, Verbrugh H. Nasal carriage of *Staphylococcus aureus*: Epidemiology, underlying mechanisms, and associated risks. Clinical Microbiology Reviews. 1997; 10(3): 505-520.

Kober M, Abreu M, Bogo M, Ferreira C, Oliveira S. Differentiation of *Salmonella enteritidis* isolates by fluorescent amplified fragment length polymorphism. Foodborne Pathogens and Disease. 2011; 8(1): 19-26.

Kohler C, von Eiff C, Peters G, Proctor R, Hecker M, Engelmann S. Physiological characterization of a heme-deficient mutant of *Staphylococcus aureus* by a proteomic approach. Journal of Bacteriology. 2003; 185(23): 6928-6937.

Kondo Y, Ito T, Ma X, Watanabe S, Kreiswirth B, Etienne J, Hiramatsu K. Combination of multiplex PCRs for staphylococcal cassette chromosome *mec* type assignment: rapid identification system for *mec*, *ccr*, and major differences in junkyard regions. Antimicrobial Agents and Chemotherapy. 2007; 51(1): 264-274.

Koreen L, Ramaswamy S, Graviss E, Naidich S, Musser J, Kreiswirth B. *spa* typing method for discriminating among *Staphylococcus aureus* isolates: implications for use of a single marker to detect genetic micro- and macrovariation. Journal of Clinical Microbiology. 2004; 42(2): 792-799.

Köser C, Holden M, Ellington M, Cartwright E, Brown N, Ogilvy-Stuart A, Hsu L, Chewapreecha C, Croucher N, Harris S, Sanders M, Enright M, Dougan G, Bentley S, Parkhill J, Fraser L, Betley J, Schulz-Trieglaff O, Smith G, Peacock S. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. The New England Journal of Medicine. 2012; 366(24): 2267-2275.

Kreisel K, Stine O, Johnson J, Perencevich E, Shardell M, Lesse A, Gordin F, Climo M, Roghmann M. USA300 methicillin-resistant *Staphylococcus aureus* bacteremia and the risk of severe sepsis : is USA300 methicillin-resistant *Staphylococcus aureus* associated with more severe infections? Diagnostic Microbiology and Infectious Disease. 2011; 70(30): 285-290.

Kreiswirth B, Kornblum J, Arbeit R, Eisner W, Maslow J, McGeer A, Low D, Novick R. Evidence for a clonal origin of methicillin resistance in *Staphylococcus aureus*. Science. 1993; 259(5092): 227-230.

Kreiswirth B. Genetics and expression of toxic shock syndrome toxin 1: overview. Reviews of Infectious Diseases. 1989; 11(Supplement 1): s97-s100.

Krziwanek K, Metz-Gercek S, Mittermayer H. Methicillin-Resistant *Staphylococcus aureus* ST398 from human patients, upper Austria. Emerging Infectious Diseases. 2009; 15(5): 766-769.

Kumagai Y, Konishi K, Gomi T, Yagishita H, Yajima A, Yoshikawa M. Enzymatic properties of dipeptidyl aminopeptidase IV produced by the periodontal pathogen *Porphyromonas gingivalis* and its participation in virulence. Infection and Immunity. 2000; 68(2): 716-724.

Kuroda M, Ohta T, Uchiyama I, Baba T, Yuzawa H, Kobayashi I, Cui L, Oguchi A, Aoki K, Nagai Y, Lian J, Ito T, Kanamori M, Matsumaru H, Maruyama A, Murakami H, Hosoyama A, Mizutani-ui Y, Takahashi N, Sawano T, Inoue R, Kaito C, Sekimizu K, Hirakawa H, Kuhara S, Goto S, Yabuzaki J, Kanehisa M, Yamashita A, Oshima K, Furuya K, Yoshino C, Shiba T, Hattori M, Ogasawara N, Hayashi H, Hiramatsu K. Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*. The Lancet. 2001; 357(9264): 1225-1240.

Kwon A, Park G, Ryu S, Lim D, Lim D, Choi C, Park Y, Lim Y. Higher biofilm formation in multidrug-resistant clinical isolates of *Staphylococcus aureus*. International Journal of Antimicrobial Agents. 2008; 32: 68-72.

Labandeira-Rey M, Couzon F, Boisset S, Brown E, Bes M, Benito Y, Barbu E, Vazquez V, Höök M, Etienne J, Vandenesch F, Bowden M. *Staphylococcus aureus* Panton-Valentine leukocidin causes necrotizing pneumonia. Science. 2007; 315(5815): 1130-1133.

Lacey R, Grinsted J. Genetic analysis of methicillin-resistant strains of *Staphylococcus aureus*; evidence for their evolution from a single clone. Journal of Medical Microbiology. 1973; 6(4): 511-526.

Ladhani S. Understanding the mechanism of action of the exfoliative toxins of *Staphylococcus aureus*. Federation of European Microbiological Societies Immunology and Medical Microbiology. 2003; 39(2): 181-189.

Lane D, Pace B, Olsen G, Stahl D, Sogin M, Pace N. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proceedings of the National Academy of Sciences of the United States of America. 1985; 82(20): 6955-6959.

Lane D. 16S/23S rRNA sequencing. Stakebrandt E, Goodfellow M, editors. Nucleic acid techniques in bacterial systematics. Chichester: John Wiley & Sons; 1991.

Lappin E, Ferguson A. Gram-positive toxic shock syndromes. The Lancet Infectious Diseases. 2009; 9(5): 281-290.

Lee L, Höök M, Haviland D, Wetsel R, Yonter E, Syribeys P, Vernchio J, Brown E. Inhibition of complement activation by a secreted *Staphylococcus aureus* protein. The Journal of Infectious Diseases. 2004; 190(3): 571-579.

Lee L, Liang X, Höök M, Brown E. Identification and characterization of the C3 binding domain of the *Staphylococcus aureus* extracellular fibrinogen-binding protein (Efb). The Journal of Biological Chemistry. 2004; 279(49): 50710-50716.

Letertre C, Perelle S, Dilasser F, Fach P. Identification of a new putative enterotoxin SEU encoded by the *egc* cluster of *Staphylococcus aureus*. Journal of Applied Microbiology. 2003; 95(1): 38-43.

Li S, Skov R, Han X, Larsen A, Larsen J, Sørum M, Wulf M, Voss A, Hiramatsu K, Ito T. Novel types of staphylococcal cassette chromosome *mec* elements identified in clonal complex 398 methicillin-resistant *Staphylococcus aureus* strains. Antimicrobial Agents and Chemotherapy. 2011; 55(6): 3046-3050.

Liang X, Ji Y. Alpha-toxin interferes with integrin-mediated adhesion and internalization of *Staphylococcus aureus* by epithelial cells. Cellular Microbiology. 2006; 8(10): 1656-1668.

Liang X, Yu C, Sun J, Liu H, Landwehr C, Holmes D, Ji Y. Inactivation of a two-component signal transduction system, SaeRS, eliminates adherence and attenuates virulence of *Staphylococcus aureus*. Infection and Immunity. 2006; 74(8): 4655-4665.

Lin Y, St Maurice M. The structure of allophanate hydrolase from *Granulibacter bethesdensis* provides insights into substrate specificity in the amidase signature family. Biochemistry. 2013; 52(4): 690-700.

Lindsay A, Holden M. *Staphylococcus aureus*: superbug, super genome? Trends in Microbiology. 2004; 12(8): 378-385.

Lindsay J, Holden M, Enright M, Francois P, Schrenzel J, McNamara P, Fournier B, Horsburgh M, Pinho M, Queck S, Otto M, Fitzgerald J, Penadés J. *Staphylococcus* molecular genetics Norfolk: Caister academic press; 2008.

Lindsay J, Knight G, Budd E, McCarthy A. Shuffling of mobile genetic elements (MGEs) in successful healthcare-associated MRSA (HA-MRSA). Mobile genetic elements. 2012; 2(5): 239-243.

Lindsay J, Moore C, Day N, Peacock S, Witney A, Stabler R, Husain S, Butcher P, Hinds J. Microarrays reveal that each of the ten dominant lineages of *Staphylococcus aureus* has a unique combination of surface-associated and regulatory genes. Journal of Bacteriology. 2006; 188(2): 669-676.

Lindsay J, Ruzin A, Ross H, Kurepina N, Novick R. The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus*. Molecular Microbiology. 1998; 29(2): 527-543.

Liu M, Liu J, Guo Y, Zhang Z. Characterization of virulence factors and genetic background of *Staphylococcus aureus* isolated from Peking university people's hospital between 2005 and 2009. Current Microbiology. 2010; 61(5): 435-443.

Liu Z, Tan C, Guo X, Kao Y, Li J, Wang L, Sancar A, Zhong D. Dynamics and mechanism of cyclobutane pyrimidine dimer repair by DNA photolyase. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108(36): 14831-14836.

Long K, Poehlsgaard J, Kehrenberg C, Schwarz S, Vester B. The Cfr rRNA methyltransferase confers resistance to phenicols, lincosamides, oxazolidinones, pleuromutilins, and streptogramin A antibiotics. Antimicrobial Agents and Chemotherapy. 2006; 50(7): 2500-2505.

Lowy F. Antimicrobial resistance : the example of *Staphylococcus aureus*. The Journal of Clinical Investigation. 2003; 111(9): 1265-1273.

Lowy F. Is *Staphylococcus aureus* an intracellular pathogen? Trends in Microbiology. 2000; 8(8): 341-343.

Luong T, Ouyang S, Bush K, Lee C. Type 1 capsule genes of *Staphylococcus aureus* are carried in a staphylococcal cassette chromosome genetic element. Journal of Bacteriology. 2002; 184(13): 3623-3629.

Lyer L, Makarova K, Koonin E, Aravind L. Comparative genomics of the Ftsk-HerA superfamily of pumping ATPases: implications for the origins of chromosomal segregation, cell division and viral capsid packaging. Nucleic Acids Research. 2004; 32(17): 5260-5279.

Ma X, Ito T, Tiensasitorn C, Jamklang M, Chongtrakoool P, Boyle-Vavra S, Daum R, Hiramatsu K. Novel type of staphylococcal cassette chromosome *mec* identified in community-acquired methicillin-resistant *Staphylococcus aureus* strains. Antimicrobial Agents and Chemotherapy 2002; 46(4): 1147-1152.

Macdonald A, Smith G. The staphylococci. In Proceedings of the Alexander Ogston centennial conferences; 1981; Aberdeen. 33-62.

Mahillon J, Chandler M. Insertion sequences. Microbiology and Molecular Biology Reviews. 1998; 62(3): 725-774.

Maiques E, Úbeda C, Tormo M, Ferrer M, Lasa Í, Novick R, Penadés J. Role of staphylococcal phage and SaPI integrase in intra- and interspecies SaPI transfer. Journal of Bacteriology. 2007; 189(15): 5608-5616.

Majerczyk C, Dunman P, Luong T, Lee C, Sadykov M, Somerville G, Bodi K, Sonenshein A. Direct targets of CodY in *Staphylococcus aureus*. Journal of Bacteriology. 2010; 192(11): 2861-2877.

Malouin F, Bryan L. Modification of penicillin-binding proteins as mechanisms of β-lactam resistance. Antimicrobial Agents and Chemotherapy. 1986; 30(1): 1-5.

Mammina C, Calà C, Bonura C, Di Carlo P, Aleo A, Fasciana T, Giammanco A, EPI-MRSA Working Group. Polyclonal non multiresistant methicillin resistant *Staphylococcus aureus* isolates from clinical cases of infection occurring in Palermo, Italy, during a one-year surveillance period. Annals of Clinical Microbiology and Antimicrobials. 2012; 11(17).

Mammina C, Calà C, Plano M, Bonura C, Vella A, Monastero R, Palma D. Ventilator-associated pneumonia and MRSA ST398, Italy. Emerging Infectious Diseases. 2010; 16(4): 730-731.

Marceau A. Functions of single-strand DNA-binding proteins in DNA replication, recombination, and repair. Methods in Molecular Biology. 2012; 922: 1-21.

Marples R, Richardson J, de Saxe M. Bacteriological characters of strains of *Staphylococcus aureus* submitted to a reference laboratory related to methicillin resistance. The Journal of Hygiene. 1986; 96(2): 217-223.

Martin V, Hope R, Livermore D, Reynolds R. BSAC Bacteraemia resistance surveillance update 2010. Poster presented in 2011; Manchester: Federation of Infection Societies Annual Conference 2011.

Martínez J. Bottlenecks in the transferability of antibiotic resistance from natural ecosystems to human bacterial pathogens. Frontiers in Microbiology. 2011; 3(2): 265.

Martinez J. The role of natural environments in the evolution of resistance traits in pathogenic bacteria. Proceedings of the Royal Society of Biological Sciences. 2009; 276(1667): 2521-2530.

Masoud W, Vogensen F, Lillevang S, Abu Al-Soud W, Sørensen S, Jakobsen M. The fate of indigenous microbiota, starter cultures, *Escherichia coli*, *Listeria innocua* and *Staphylococcus aureus* in Danish raw milk and cheeses determined by pyrosequencing and quantitative real time (qRT)-PCR. International Journal of Food Microbiology. 2012; 153(1-2): 192-202.

Massimi I, Park E, Rice K, Müller-Ester W, Sauder D, McGavin M. Identification of a novel maturation mechanism and restricted substrate specificity for the SspB cyteine protease of *Staphylococcus aureus*. The Journal of Biological Chemistry. 2002; 277(44): 41770-41777.

McAleese F, Wu S, Sieradzki K, Dunman P, Murphy E, Projan S, Tomasz A. Overexpression of genes of the cell wall stimulon in clinical isolates of *Staphylococcus aureus* exhibiting vancomycin-intermediate-S. *aureus*-type resistance to vancomycin. Journal of Bacteriology. 2006; 188(3): 1120-1133.

McCarthy A, Lindsay J. Genetic variation in *Staphylococcus aureus* surface and immune evasion genes is lineage associated: implications for vaccine design and host-pathogen interactions. BioMed Central Microbiology. 2010; 10.

McClintock B. The origin and behavoir of mutable loci in maize. Proceedings of the National Academy of Sciences of the United States of America. 1950; 36(6): 344-355.

McClure J, Conly J, Elsayed S, Zhang K. Multiplex PCR assay to facilitate identification of the recently described staphylococcal cassette chromosome *mec* type VIII. Molecular and Cellular Probes. 2010; 24(4): 229-232.

McDougal L, Steward C, Killgore G, Chaitram J, Mcallister S, Tenover F. Pulsed-field gel electrophoresis typing of oxacillin-resistant *Staphylococcus aureus* isolates from the United States : establishing a national database. Journal of Clinical Microbiology. 2003; 41(11): 5113-5120.

Melles D, Gorkink R, Boelens H, Snijders S, Peeters J, Moorhouse M, van der Spek P, van Leeuwen W, Simons G, Verbrugh H, van Belkum A. Natural population dynamics and expansion of pathogenic clones of *Staphylococcus aureus*. The Journal of Clinical Investigation. 2004; 114(12): 1732-1740.

Melles D, Schouls L, François P, Herzig S, Verbrugh H, van Belkum A, Schrenzel J. High-throughput typing of *Staphylococcus aureus* by amplified fragment length polymorphism (AFLP) or multi-locus variable number of tandem repeat analysis (MLVA) reveals consistent strain relatedness. European Journal of Clinical Microbiology and Infectious Diseases. 2009; 28(1): 39-45.

Melles D, Tenover F, Kuehnert M, Witsenboer H, Peeters J, Verbrugh H, van Belkum A. Overlapping population structures of nasal isolates of *Staphylococcus aureus* from healthy Dutch and American individual. Journal of Clinical Microbiology. 2008; 46(1): 235-241.

Melles D, van Leeuwen W, Snijders S, Horst-Kreft D, Peeters J, Verbrugh H, van Belkum A. Comparison of multilocus sequence typing (MLST), pulsed-field gel electrophoresis (PFGE), and amplified fragment length polymorphism (AFLP) for genetic typing of *Staphylococcus aureus*. Journal of Microbiological Methods. 2007; 69(2): 371-375.

Milheiriço C, Oliveira D, de Lencastre H. Multiplex PCR strategy for subtyping the staphylococcal cassette chromosome *mec* type IV in methicillin-resistant *Staphylococcus aureus*: 'SCC*mec* IV multiplex'. Journal of Antimicrobial Chemotherapy. 2007; 60(1): 42-48.

Milheiriço C, Oliveira D, de Lencastre H. Update to the multiplex PCR strategy for assignment of *mec* element types in *Staphylococcus aureus*. Antimicrobial Agents and Chemotherapy. 2007; 51(9): 3374-3377.

Milheiriço C, Portelinha A, Krippahl L, de Lencastre H, Oliveira D. Evidence for a purifying selection acting on the β-lactamase locus in epidemic clones of methicillin-resistant *Staphylococcus aureus*. BioMed Central Microbiology. 2011; 11(76).

Mira A, Martín-Cuadrado A, D'Auria G, Rodríguez-Valera F. The bacterial pan-genome: a new paradigm in microbiology. International Microbiology. 2010; 13(2): 45-57.

Moise P, Smyth D, Robinson D, El-Fawal N, McCalla C, Sakoulas G. Genotypic and phenotypic relationships among methicillin-resistant *Staphylococcus aureus* from three multicentre bacteraemia studies. Journal of Antimicrobial Chemotherapy. 2009; 63(5): 873-876.

Monecke S, Coombs G, Shore A, Coleman D, Akpaka P, Borg M, Chow H, Ip M, Jatzwauk L, Jonas D, Kadlec K, Kearns A, Laurent F, O'Brien F, Pearson J, Ruppelt A, Schwarz S, Scicluna E, Slickers P, Tan H, Weber S, Ehricht, R. A field guide to pandemic, epidemic and sporadic clones of methicillin-resistant *Staphylococcus aureus*. Public Library of Science One. 2011; 6(4): e17936.

Mostofsky E, Lipsitch M, Regev-Yochay G. Is methicillin-resistant *Staphylococcus aureus* replacing methicillin-susceptible *S. aureus*? Journal of Antimicrobial Chemotherapy. 2011; 66(10): 2199-2214.

Murray B. Vancomycin-resistant enterococcal infections. The New England Journal of Medicine. 2000; 342(10): 710-721.

Musser J, Kapur V. Clonal analysis of methicillin-resistant *Staphylococcus aureus* strains from intercontinental sources: association of the *mec* gene with divergent phylogenetic lineages implies dissemination by horizontal transfer and recombination. Journal of Clinical Microbiology. 1992; 30(8): 2058-2063.

Nakaminami H, Noguchi N, Nishijima S, Kurokawa I, Sasatsu M. Characterization of the pTZ2162 encoding multidrug efflux gene *qacB* from *Staphylococcus aureus*. Plasmid. 2008; 60(2): 108-117.

Nakamura Y, Koyama K, Matsushima M. VNTR (variable number of tandem repeat) sequences as transcriptional, translational, or functional regulators. Journal of Human Genetics. 1998; 43(3): 149-152.

Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M, Martin C, Fujimoto E, Hoff M, Kumlin E, White R. Variable number of tandem repeat (VNTR) markers for human gene mapping. Science. 1987; 235(4796): 1616-1622.

Nakano M, Zhu Y, Haga K, Yoshikawa H, Sonenshein A, Zuber P. A mutation in the 3-phosphoglycerate kinase gene allows anaerobic growth of *Bacillus subtilis* in the absence of ResE kinase. Journal of Bacteriology. 1999; 181(22): 7087-7097.

Navaratna M, Sahl H, Tagg J. Identification of genes encoding two-component lantibiotic production in *Staphylococcus aureus* C55 and other phage group II *S. aureus* strains and demonstration of an association with the exfoliative toxin B gene. Infection and Immunity. 1999; 67(8): 4268-4271.

Noble W, Virani Z, Cree R. Co-transfer of vancomycin and other resistance genes from *Enterococcus faecalis* NCTC 12201 to *Staphylococcus aureus*. Federation of European Microbiological Societies Microbiology Letters. 1992; 72(2): 195-198.

Noto M, Kreiswirth B, Monk A, Archer G. Gene acquisition at the insertion site for SCC*mec*, the genomic island conferring methicillin resistance in *Staphylococcus aureus*. Journal of Bacteriology. 2008; 190(4): 1276-1283.

Nouwen J, Boelens H, van Belkum A, Verbrugh H. Human factor in *Staphylococcus aureus* nasal carriage. Infection and Immunity. 2004; 72(11): 6685-6688.

Nouwen J, Fieren M, Snijders S, Verbrugh H, van Belkum A. Persistent(not intermittent) nasal carriage of *Staphylococcus aureus* is the determinant of CPD-related infections. Kidney International. 2005; 67: 1084-1092.

Nübel U, Roumagnac P, Feldkamp M, Song J, Ko K, Huang Y, Coombs G, Ip M, Westh H, Skov R, Struelens M, Goering R, Strommenger B, Weller A, Witte W, Achtman M. Frequent emergence and limited geographic dispersal of methicillin-resistant *Staphylococcus aureus*. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105(37): 14130-14135.

Ogston A. Micrococcus poisoning. The Lancet. 1880;17(Part 1): 24-58.

Ohki R, Giyanto , Tateno K, Masuyama W, Moriya S, Kobayashi K, Ogasawara N. The BceRS two-component regulatory system induces expression of the bacitracin transporter, BceAB, in *Bacillus subtilis*. Molecular Biology. 2003; 49(4): 1135-1144.

Oliveira D, de Lencastre H. Methicillin-resistance in *Staphylococcus aureus* is not affected by the overexpression in trans of the *mecA* gene repressor: a surprising observation. Public Library of Science One. 2011; 6(8): e232827.

Oliveira D, de Lencastre H. Multiplex PCR strategy for rapid identification of structural types and variants of the *mec* element in methicillin-resistant *Staphylococcus aureus*. Antimicrobial Agents and Chemotherapy. 2002; 46(7): 2155-2161.

Oliveira D, Milheiriço C, de Lencastre H. Redefining a structural variant of staphylococcal cassette chromosome *mec*, SCC*mec* type VI. Antimicrobial Agents and Chemotherapy. 2006; 50(10): 3457-3459.

Oliveira D, Tomasz A, de Lencastre H. The evolution of pandemic clones of methicillin-resistant *Staphylococcus aureus*: identification of two ancestral genetic backgrounds and the associated *mec* elements. Microbial Drug Resistance. 2001; 7(4): 349-361.

Omoe K, Hu D, Takahashi-Omoe H, Nakane A, Shinagawa K. Identification and characterization of a new staphylococcal enterotoxin-related putative toxin encoded by two kinds of plasmids. Infection and Immunity. 2003; 71(10): 6088-6094.

Orwin P, Fitzgerald J, Leung D, Gutierrez J, Bohach G, Schlievert P. Characterization of *Staphylococcus aureus* enterotoxin L. Infection and Immunity. 2003; 71(5): 2916-2919.

Orwin P, Leung D, Donahue H, Novick R, Schlievert P. Biochemical and biological properties of Staphylococcal enterotoxin K. Infection and Immunity. 2001; 69(1): 360-366.

Otto M, Peschel A, Götz F. Producer self-protection against the lantibiotic epidermin by the ABC transporter EpiFEG of *Staphylococcus epidermidis* Tü3298. Federation of European Microbiological Societies Microbiology Letters. 1998; 166(2): 203-211.

Palma M, Haggar A, Flock J. Adherence of *Staphylococcus aureus* is enhanced by an endogenous secreted protein with broad binding activity. Journal of Bacteriology. 1999; 181(9): 2840-2845.

Palmqvist N, Patti J, Tarkowski A, Josefsson E. Expression of staphylococcal clumping factor A impedes macrophage phagocytosis. Microbes and Infection. 2004; 6(2): 188-195.

Patton T, Rice K, Foster M, Bayles K. The *Staphylococcus aureus cidC* gene encodes a pyruvate oxidase that affects acetate metabolism and cell death in stationary phase. Molecular Microbiology. 2005; 56(6): 1664-1674.

Peacock S, de Silva G, Justice A, Cowland A, Moore C, Winearls C, Day N. Comparison of multilocus sequence typing and pulsed-field gel electrophoresis as tools for typing *Staphylococcus aureus* isolates in a microepidemiological setting. Journal of Clinical Microbiology. 2002; 40(10): 3764-3770.

Pearman J, Coombs G, Grubb W, O'Brien F. A British epidemic strain of methicillin-resistant *Staphylococcus aureus* (UK EMRSA-15) in Western Australia. The Medical Journal of Australia. 2001; 174(12): 662.

Pearson A, Chronias A, Murray M. Voluntary and mandatory surveillance for methicillin-resistant *Staphylococcus aureus*, (MRSA) and methicillin-susceptible *S. aureus* (MSSA) bacteraemia in England. Journal of Antimicrobial Chemotherapy. 2009; 64(1): 11-17.

Peck K, Baek J, Song J, Ko K. Comparison of genotypes and enterotoxin genes between *Staphylococcus aureus* isolates from blood and nasal colonizers in a Korean hospital. Journal of Korean Medical Science. 2009; 24(4): 585-591.

Périchon B, Courvalin P. VanA-type vancomycin-resistant *Staphylococcus aureus*. Antimicrobial Agents and Chemotherapy. 2009; 53(11): 4580-4587.

Platt S, Pichon B, George R, Green J. A bioinformatics pipeline for high-throughput microbial multilocus sequence typing (MLST) analyses. Clinical Microbiology and Infection. 2006; 12(11): 1144-1146.

Pohl K, Francois P, Stenz L, Schlink F, Geiger T, Herbert S, Goerke C, Schrenzel J, Wolz C. CodY in *Staphylococcus aureus*: a regulatory link between metabolism and virulence gene expression. Journal of Bacteriology. 2009; 191(9): 2953-2963.

Poliakov A, Chang J, Spilman M, Damle P, Christie G, Mobley J, Dokland T. Capsid size determination by *Staphylococcus aureus* pathogenicity island SaPI1 involves specific incorporation of SaPI1 proteins into procapsids. Journal of Molecular Biology. 2008; 380(3): 465-475.

Potempa J, Fedak D, Dubin A, Mast A, Travis J. Proteolytic inactivation of α-1-anti-chymotrypsin. The Journal of Biological Chemistry. 1991; 266(32): 21482-21487.

Potempa J, Watorek W, Travis J. The inactivation of human plasma α1-proteinase inhibitor by proteinases from *Staphylococcus aureus*. The Journal of Biological Chemistry. 1986; 261(30): 14330-14334.

Prevost G, Couppie P, Prevost P, Gayet S, Petiau P, Cribier B, Monteil H, Piemont Y. Epidemiological data on *Staphylococcus aureus* strains producing synergohymenotropic toxins. Journal of Medical Microbiology. 1995; 42(4): 237-245.

Prevost G, Jaulhac B, Piemont Y. DNA fingerprinting by pulsed-field gel electrophoresis is more effective than ribotyping in distinguishing among methicillin-resistant *Staphylococcus aureus* isolate. Journal of Clinical Microbiology. 1992; 30(4): 967-973.

Projan S, Novick R. The molecular basis of pathogenicity. Crossley K, Archer G, editors. The staphylococci in human disease. New York: Churchill Livingstone; 1997.

Pulsen I, Reizer J, Jin R, Lin E, Saier M. Functional genomic studies of dihydroxyacetone utilization in *Escherichia coli*. Microbiology. 2000; 146(Part 10): 2343-2344.

Qi W, Ender M, O'Brien F, Imhof A, Ruef C, McCallum N, Berger-Bächi B. Molecular epidemiology of methicillin-resistant *Staphylococcus aureus* in Zürich, Switzerland (2003): prevalence of type IV SCC*mec* and a new SCC*mec* element associated with isolates from intravenous drug users. Journal of Clinical Microbiology. 2005; 43(10): 5164-5170.

Rahman A, Nariya H, Izaki K, Kato I, Kamio Y. Molecular cloning and nucleotide sequence of leukocidin F-component gene (*lukF*) from methicillin resistant *Staphylococcus aureus*. Biochemical and Biophysical Research Communications. 1992; 184(2): 640-646.

Ramsey M, Bradley S, Kauffman C, Morton T. Identification of chromosomal location of *mupA* gene, encoding lov-level mupirocin resistance in staphylococcal isolates. Antimicrobial Agents and Chemotherapy. 1996; 40(12): 2820-2823.

Rao G, Kearns A, Edwards G. Decline and fall of epidemic methicillin-resistant *Staphylococcus aureus*-16. Journal of Hospital Infection. 2011; 79(3): 269-270.

Raux E, Schubert H, Warren M. Biosynthesis of cobalamin (vitamin B12): a bacterial conundrum. Cellular and Molecular Life Sciences. 2000; 57(13-14): 1880-1893.

Recsei P, Kreiswirth B, O'Reilly M, Schlievert P, Gruss A, Novick R. Regulation of exoprotein gene expression in *Staphylococcus aureus* by *agr*. Molecular and General Genetics. 1986; 202(1): 58-61.

Redfield R. Do bacteria have sex? Nature Reviews Genetics. 2001; 2(8): 634-639.

Resende C, Figueiredo A. Discrimination of methicillin-resistant *Staphylococcus aureus* from borderline-resistant and susceptible isolates by different methods. Journal of Medical Microbiology. 1997 ; 46(2): 145-149.

Reyes J, Rincón S, Díaz L, Panesso D, Contreras G, Zurita J, Carrillo C, Rizzi A, Guzmán M, Adachi J, Chowdhury S, Murray B, Arias C. Dissemination of methicillin-resistant *Staphylococcus aureus* (MRSA), USA300 sequence type 8 lineage in Latin-America. Clinical Infectious Diseases. 2009; 49(12): 1861-1867.

Reynolds R, Hope R, Williams L, BSAC Working Parties on Resistance Surveillance. Survey, laboratory and statistical methods for the BSAC Resistance Surveillance Programmes. Journal of Antimicrobial Chemotherapy. 2008; 62(Supplement 2): ii15-ii28.

Reynolds R. Antimicrobial resistance in the UK and Ireland. Journal of Antimicrobial Chemotherapy. 2009; 64(Supplement 1): i19-i23.

Richardson A, Libby S, Fang F. A nitric oxide-inducible lactate dehydrogenase enables *Staphylococcus aureus* to resist innate immunity. Science. 2008; 319(5870): 1672-1676.

Richardson J, Reith S. Characterization of a strain of methicillin-resistant *Staphylococcus aureus* (EMRSA-15) by conventional and molecular methods. Journal of Hospital Infection. 1993; 25(1): 45-52.

Ridom GmbH. Ridom SpaServer. [Online]; 2012 [cited 2012]. Available from: www.spaserver.ridom.de.

Roberts A, Withers P. statistiXL. [Online]; 2009 [cited 2012]. Available from: http://www.statistixl.com/about/about.aspx.

Robinson D, Enright M. Evolution of *Staphylococcus aureus* by large chromosomal replacements. Journal of Bacteriology. 2004; 186(4): 1060-1064.

Robinson D, Enright M. Evolutionary models of the emergence of methicillin-resistant *Staphylococcus aureus*. Antimicrobial Agents and Chemotherapy. 2003; 47(12): 3926-3934.

Ronaghi M, Uhlén M, Nyrén P. A sequencing method based on real-time pyrophosphate. Science. 1998; 281(5375): 363-365.

Ronaghi M. Pyrosequencing sheds light on DNA sequencing. Genome Research. 2001; 11(1): 3-11.

Rooijakkers S, Ruyken M, Roos A, Daha M, Presanis J, Sim R, van Wamel W, van Kessel K, van Strijp J. Immune evasion by a staphylococcal complement inhibitor that acts on C3 convertases. Nature Immunology. 2005; 6(9): 920-927.

Rooijakkers S, van Wamel W, Ruyken M, van Kessel K, van Strijp J. Anti-opsonic properties of staphylokinase. Microbes and Infection. 2005; 7(3): 476-484.

Roos T, van Passel M. A quantitative account of genomic island acquisitions in prokaryotes. BioMed Central Genomics. 2011; 12.

Rosenbach A. Mikro-Qrganismen bei den Wund-Infections-Krankheiten des Menschen. 1884: 18.

Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. Misener S, Krawetz S, editors. Bioinformatics Methods and Protocols. New York: Springer-Verlag; 1999.

Sakwinska O, Kuhn G, Balmelli C, Francioli P, Giddey M, Perreten V, Riesen A, Zysset F, Blanc D, Moreillon P. Genetic diversity and ecological success of *Staphylococcus aureus*. Applied and Environmental Microbiology. 2009; 75(1): 175-183.

Sanger F, Coulson A. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. Journal of Molecular Biology. 1975; 94(3): 441-448.

Sanger F, Nicklen S, Coulson A. DNA sequencing with chain-terminating inhibitors. Proceedings for the National Academy of Sciences of the United States of America. 1977; 74(12): 5463-5467.

Saunders N, Underwood A, Kearns A, Hallas G. A virulence-associated gene microarray: a tool for investigation of the evolution and pathogenic potential of *Staphylococcus aureus*. Microbiology. 2004; 150(Part 11): 3763-3771.

Schlag S, Fuchs S, Nerz C, Gaupp R, Engelmann S, Liebeke M, Lalk M, Hecker M, Götz F. Characterization of the oxygen-responsive NreABC regulon of *Staphylococcus aureus*. Journal of Bacteriology. 2008; 190(23): 7847-7858.

Schlievert P, Strandberg K, Lin Y, Peterson M, Leung D. Secreted virulence factor comparison between methicillin-resistant and methicillin-sensitive *Staphylococcus aureus*, and its relevance to atopic dermatitis. The Journal of Allergy and Clinical Immunology. 2010; 125(1): 39-49.

Schnaith A, Kashkar H, Leggio S, Addicks K, Krönke M, Krut O. *Staphylococcus aureus* subvert autophagy for induction of caspase-independent host cell death. The Journal of Biological Chemistry. 2007; 282(4): 2695-2706.

Schwan W, Langhorne M, Ritchie H, Stover C. Loss of hemolysin expression in *Staphylococcus aureus agr* mutants correlates with selective survival during mixed infections in murine abscesses and wounds. Federation of European Microbiological Societies Immunology and Medical Microbiology. 2003; 38(1): 23-28.

Schwartz D, Cantor C. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell. 1984; 37(1): 67-75.

Schwarz-Linek U, Werner J, Pickford A, Gurusiddappa S, Kim J, Pilka E, Brigg J, Gough T, Höök M, Campbell I, Potts J. Pathogenic bacteria attach to human fibronectin through a tandem β-zipper. Nature. 2003; 423(6963): 177-181.

Selander R, Caugant D, Ochman H, Musser J, Gilmour M, Whittam T. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. Applied and Environmental Microbiology. 1986; 51(5): 873-884.

Severin A, Tabei K, Tenover F, Chung M, Clarke N, Tomasz A. High level oxacillin and vancomycin resistance and altered cell wall composition in *Staphylococcus aureus* carrying the staphylococcal *mecA* and the enterococcal *vanA* gene complex. The Journal of Biological Chemistry. 2004; 279(5): 3398-3407.

Shockman G, Barrett J. Structure, function, and assembly of cell walls of Gram-positive bacteria. Annual Reviews of Microbiology. 1983; 37: 501-527.

Shopsin B, Gomez M, Montgomery S, Smith D, Waddington M, Dodge D, Bost D, Riehman M, Naidich S, Kreiswirth B. Evaluation of protein A gene polymorphic region DNA sequencing for typing of *Staphylococcus aureus* strains. Journal of Clinical Microbiology. 1999; 37(11): 3556-3563.

Shore A, Deasy E, Slickers P, Brennan G, O'Connell B, Monecke S, Ehricht R, Coleman D. Detection of staphylococcal cassette chromosome *mec* type XI carrying highly divergent *mecA*, *mecI*, *mecR1*, *blaZ*, and *ccr* genes in human clinical isolates of clonal complex 130 methicillin-resistant *Staphylococcus aureus*. Antimicrobial Agents and Chemotherapy. 2011; 55(8): 3765-3773.

Shore A, Rossney A, Keane C, Enright M, Coleman D. Seven novel variants of the staphylococcal chromosomal cassette *mec* in methicillin-resistant *Staphylococcus aureus* isolates from Ireland. Antimicrobial Agents and Chemotherapy. 2005; 49(5): 2070-2083.

Shore A, Rossney A, Kinnevey P, Brennan O, Creamer E, Sherlock O, Dolan A, Cunney R, Sullivan D, Goering R, Humphreys H, Coleman D. Enhanced discrimination of highly clonal ST22-methicillin-resistant *Staphylococcus aureus* IV isolates achieved by combining *spa*, *dru*, and pulsed-field gel electrophoresis typing data. Journal of Clinical Microbiology. 2010; 48(5): 1839-1852.

Shore A, Rossney A, O'Connell B, Herra C, Sullivan D, Humphreys H, Coleman D. Detection of staphylococcal cassette chromosome *mec*-associated DNA segments in multiresistant methicillin-susceptible *Staphylococcus aureus* (MSSA) and identification of *Staphylococcus* epidermidis *ccrAB4* in both methicillin-resistant *S. aureus* and MSSA. Antimicrobial Agents and Chemotherapy. 2008; 52(12): 4407-4419.

Shukla S, Kislow J, Briska A, Henkhaus J, Dykes C. Optical mapping reveals a large genetic inversion between two methicillin-resistant *Staphylococcus aureus* strains. Journal of Bacteriology. 2009; 191(18): 5717-5723.

Shukla S, Patrangi M, Stahl B, Briska A, Stemper M, Wagner T, Zentz E, Callister S, Lovrich S, Henkhaus J, Dykes C. Comparative whole-genome mapping to determine *Staphylococcus aureus* genome size, virulence motifs, and clonality. Journal of Clinical Microbiology. 2012; 50(11): 3526-3533.

Sieradzki K, Chung M, Tomasz A. Role of a sodium-dependent symporter homologue in the thermosensitivity of β-lactam antibiotic resistance and cell wall composition in *Staphylococcus aureus*. Antimicrobial Agents and Chemotherapy. 2008; 52(2): 505-512.

Siguier P, Filée J, Chandler M. Insertion sequences in prokaryotic genomes. Current Opinion in Microbiology. 2006; 9(5): 526-531.

Silvestri L, Hill L. Agreement between deoxyribonucleic acid base composition and taxometric classification of Gram-positive cocci. Journal of Bacteriology. 1965; 90(1): 136-140.

Simpson E. Measurement of diversity. Nature. 1949; 163(4148): 688.

Singleton P. Bacteria in biology, biotechnology and medicine Chichester: John Wiley & Sons Ltd; 2004.

Sinha B, François P, Nüße O, Foti M, Hartford O, Vaudaux P, Foster T, Lew D, Herrmann M, Krause K. Fibronectin-binding protein acts as *Staphylococcus aureus* invasin via fibronectin bridging to integrin α5β1. Cellular Microbiology. 1999; 1(2): 101-117.

Song L, Hobaugh M, Shustak C, Cheley S, Bayley H, Gouaux J. Structure of staphylococcal α-hemolysins, a heptameric transmembrane pore. Science. 1996; 274(5294): 1859-1866.

Southern E. Detection of specific sequences among DNA fragments separated by gel electrophoresis. 1975. Biotechnology. 1992; 24: 122-139.

Stackebrandt E, Woese C. A phylogenetic dissection of the family Micrococcaceae. Current Microbiology. 1979; 2: 317-322.

Stegger M, Andersen P, Kearns A, Pichon B, Holmes M, Edwards G, Laurent F, Teale C, Skov R, Larsen A. Rapid detection, differentiation and typing of methicillin-resistant *Staphylococcus aureus* harbouring either *mecA* or the new *mecA* homologue $mecA_{LGA251}$. Clinical Microbiology and Infection. 2012; 18(4): 395-400.

Stegger M, Price L, Larsen A, Gillece J, Waters A, Skov R, Andersen P. Genome sequence of *Staphylococcus aureus* strain 11819-97, an ST80-IV European community-acquired methicillin-resistant isolate. Journal of Bacteriology. 2012; 194(6): 1625-1626.

Stephens W. Pulsed mass spectrometer with time dispersion. Bulletin of the American Physical Society. 1946; 21(2): 22.

Stewart P, Costerton J. Antibiotic resistance of bacteria in biofilms. The Lancet. 2001; 358(9276): 135-138.

Symms C, Cookson B, Stanley J, Hookey J. Analysis of methicillin-resistant *Staphylococcus aureus* IS*1181* profiling. Epidemiology and Infection. 1998; 120(3): 271-279.

Sztramko R, Katz K, Antoniou T, Mulvey M, Brunetta J, Crouzat F, Kovacs C, Merkley B, Tilley D, Loutfy M. Community-associated methicillin-resistant *Staphylococcus aureus* infections in men who have sex with men: a case series. 2007; 18(4): 257-261.

Takeuchi F, Watanab S, Baba T, Yuzawa H, Ito T, Morimoto Y, Kuroda M, Cui L, Takahashi M, Ankai A, Baba S, Fukui S, Lee J, Hiramatsu K. Whole-genome sequencing of *Staphylococcus haemolyticus* uncovers the extreme plasticity of its genome and the evolution of human-colonizing staphylococcal species. Journal of Bacteriology. 2005; 187(21): 7292-7308.

Tallent S, Langston T, Moran R, Christie G. Transducing particles of *Staphylococcus aureus* pathogenicity island SaPI1 are comprised of helper phage-encoded proteins. Journal of Bacteriology. 2007; 189(20): 7520-7524.

Tenover F, Arbeit R, Archer G, Biddle J, Byrne S, Goering R, Hancock G, Hébert G, Hill B, Hollis R, Jarvis W, Kreiswirth B, Eisner W, Maslow J, McDougal L, Miller J, Mulligan M, Pfaller M. Comparison of traditional and molecular methods of typing isolates of *Staphylococcus aureus*. Journal of Clinical Microbiology. 1994; 32(2): 407-415.

Tenover F, Arbeit R, Goering R, Mickelsen P, Murray B, Persing D, Swaminathan B. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. Journal of Clinical Microbiology. 1995; 33(9): 2233-2239.

Tenover F, Goering R. Methicillin-resistant *Staphylococcus aureus* strain USA300: origin and epidemiology. Journal of Antimicrobial Chemotherapy. 2009; 64(3): 441-446.

Thakker M, Park J, Carey V, Lee J. *Staphylococcus aureus* serotype 5 capsular polysaccharide is antiphagocytic and enhances bacterial virulence in a murine bacteremia model. Infection and Immunity. 1998; 66(11): 5183-5189.

Thompson R, Cabezudo I, Wenzel R. Epidemiology of nosocomial infections caused by methicillin-resistant *Staphylococcus aureus*. Annals of Internal Medicine. 1982; 97(3): 309-317.

Thorne N, Evans J, Smith E, Hawkey P, Gharbia S, Arnold C. An IS*6110*-targeting fluorescent amplified fragment length polymorphism alternative to IS*6110* restriction fragment length polymorphism analysis for *mycobacterium tuberculosis* DNA fingerprinting. Clinical Microbiology and Infection. 2007; 13(10): 964-970.

Tien W, White D. Linear sequential arrangement of genes for the biosynthetic pathway of protoheme in *Staphylococcus aureus*. Proceedings for the National Academy of Sciences of the United States of America. 1968; 61(4): 1392-1398.

Tormo M, Ferrer M, Maiques E, Úbeda C, Selva L, Lasa Í, Calvete J, Novick R, Penadés J. *Staphylococcus aureus* pathogenicity island DNA is packaged in particles composed of phage proteins. Journal of Bacteriology. 2008; 190(7): 2434-2440.

Trindade P, McCulloch J, Oliveira G, Mamizuka E. Molecular techniques for MRSA typing: current issues and perspectives. The Brazilian Journal of Infectious Diseases. 2003; 7(1): 32-43.

Tuchscherr L, Medina E, Hussain M, Völker W, Heitmann V, Neimann S, Holzinger D, Roth J, Proctor R, Becker K, Peters G, Löffler B. *Staphylococcus aureus* phenotype switching: an effective bacterial strategy to escape host immune response and establish a chronic infection. European Molecular Biology Organization Molecular Medicine. 2011; 3(3): 129-141.

Tulinski P, Fluit A, Wagenaar J, Mevius D, van de Vijver L, Duim B. Methicillin-resistant coagulase-negative staphylococci on pig farms as a reservoir of heterogenous staphylococcal cassette chromosome *mec* elements. Applied and Environmental Microbiology. 2012; 78(2): 299-304.

Twort F. An investigation on the nature of ultra-microscopic viruses. The Lancet. 1915; 186(4814): 1241-1243.

Underwood A. ALFIE AFLP Fragment Predictor Program. [Online]; 2008 [cited 2011]. Available from: http://www.hpa-bioinformatics.org.uk/cgi-bin/ALFIE/index.cgi.

van Belkum A, Scherer S, van Alphen L, Verbrugh H. Short-sequence DNA repeats in prokaryotic genomes. Microbiology and Molecular Biology Reviews. 1998; 62(2): 275-293.

van Belkum A, Tassios P, Dijkshoorn L, Haeggman S, Cookson B, Fry N, Fussing V, Green J, Feil E, Gerner-Smidt P, Brisse S, Struelens M. Guidelines for the validation and application of typing methods for use in bacterial epidemiology. Clinical Microbiology and Infectious Diseases. 2007; 13(Supplement 3): 1-46.

van Belkum A, Verkaik N, de Vogel C, Boelens H, Verveer J, Nouwen J, Verbrugh H, Wertheim F. Reclassification of *Staphylococcus aureus* nasal carriage types. The Journal of Infectious Diseases. 2009; 199(12): 1820-1826.

van Belkum A. Short sequence repeats in microbial pathogenesis and evolution. Cellular and Molecular Life Sciences. 1999; 56(9-10): 729-734.

van Leeuwen W, Libregts C, Schalk M, Veuskens J, Verbrugh H, van Belkum A. Binary typing of *Staphylococcus aureus* strains through reversed hybridization using digoxigenin-universal linkage system-labeled bacterial genomic DNA. Journal of Clinical Microbiology. 2001; 39(1): 328-331.

van Leeuwen W, Melles D, Alaidan A, Al-Ahdal M, Boelens H, Snijders S, Wertheim H, van Duijkeren E, Peeters J, van der Spek P, Gorkink R, Simons G, Verbrugh H, van Belkum A. Host- and tissue-specific pathogenic traits of *Staphylococcus aureus*. Journal of Bacteriology. 2005; 187(13): 4584-4591.

van Leeuwen W, Verbrugh H, van der Velden J, van Leeuwen N, Heck M, van Belkum A. Validation of binary typing for *Staphylococcus aureus* strains. Journal of Clinical Microbiology. 1999; 37(3): 664-674.

van Wamel W, Rooijakkers S, Ruyken M, van Kessel K, van Strijp J. The innate immune modulators staphylococcal complement inhibitor and chemotaxis inhibitory protein of *Staphylococcus aureus* are located on β-hemolysin-converting bacteriophages. Journal of Bacteriology. 2006; 188(4): 1310-1315.

Varshney A, Mediavilla J, Robiou N, Guh A, Wang X, Gialanella P, Levi M, Kreiswirth B, Fries B. Diverse enterotoxin gene profiles among clonal complexes of *Staphylococcus aureus* isolates from the Bronx, New York. Applied and Environmental Microbiology. 2009; 75(21): 6839-6849.

Veiga H, Pinho M. Inactivation of the SauI type I restriction-modification system is not sufficient to generate *Staphylococcus aureus* strains capable of efficiently accepting foreign DNA. Applied and Environmental Microbiology. 2009; 75(10): 3034-3038.

Vela D, Guerreiro M, Fontdevila A. Adaptation of the AFLP technique as a new tool to detect genetic instability and transposition in interspecific hybrids. Biotechniques. 2011; 50(4): 247-250.

von Eiff C, Becker K, Machka K, Stammer H, Peters G. Nasal carriage as a source of *Staphylococcus aureus* bacteremia. The New England Journal of Medicine. 2001; 344(1): 11-16.

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research. 1995; 23(21): 4407-4414.

Waksman S. What is an antibiotic or an antibiotic substance? Mycological Society of America. 1947; 39(5): 565-569.

Waldron D, Lindsay J. Sau1: a novel lineage-specific type I restriction-modification system that blocks horizontal gene transfer into *Staphylococcus aureus* and between S. *aureus* isolates of different lineages. Journal of Bacteriology. 2006; 188(15): 5578-5585.

Wang L, Archer G. Roles of *ccrA* and *ccrB* in excision and integration of staphylococcal cassette chromosome *mec*, a *Staphylococcus aureus* genomic island. Journal of Bacteriology. 2010; 192(12): 3204-3212.

Wang W, Chiueh T, Sun J, Tsao S, Lu J. Molecular typing and phenotype characterization of methicillin-resistant *Staphylococcus aureus* isolates from blood in Taiwan. Public Library of Science One. 2012; 7(1): e30394.

Wannet W, Spalburg E, Heck M, Pluister G, Tiemersma E, Willems R, Huijsdens X, Neeling A, Etienne J. Emergence of virulent methicillin-resistant *Staphylococcus aureus* strains carrying Panton-Valentine leucocidin genes in the Netherlands. Journal of Clinical Microbiology. 2005; 43(7): 3341-3345.

Wardenburg J, Williams W, Missiakas D. Host defences against *Staphylococcus aureus* infection require recognition of bacterial lipoproteins. Proceedings of the National Academy of Sciences of the United States of America. 2006; 103(37): 13831-13836.

Watanabe S, Ito T, Morimoto Y, Takeuchi F, Hiramatsu K. Precise excision and self-integration of a composite transposon as a model for spontaneous large-scale chromosome inversion/deletion of the *Staphylococcus haemolyticus* clinical strain JCSC1435. Journal of Bacteriology. 2007; 189(7): 2921-2925.

Watson H, Walker N, Shaw P, Bryant T, Wendell P, Fothergill L, Perkins R, Conroy S, Dobson M, Tuite M. Sequence and structure of yeast phosphoglycerate kinase. The European Molecular Biology Organization Journal. 1982; 1(12): 1635-1640.

Weese J. Methicillin-resistant *Staphylococcus aureus* in animals. Institute for Laboratory Animal Research Journal. 2010; 51(3): 233-244.

Weigel L, Clewell D, Gill S, Clark N, McDougal L, Flannagan S, Kolonay J, Shetty J, Killgore G, Tenover F. Genetic analysis of a high-level vancomycin-resistant isolate of *Staphylococcus aureus*. Science. 2003; 302(5650): 1569-1571.

Weller T. Methicillin-resistant *Staphylococcus aureus* typing methods: which should be the international standard? Journal of Hospital Infection. 2000; 44(3): 160-172.

Wertheim H, Vos M, Ott A, van Belkum A, Voss A, Kluytmans J, van Keulen P, Vandenbroucke-Grauls C, Meester M, Verbrugh H. Risk and outcome of nosocomial *Staphylococcus aureus* bacteraemia in nasal carriers versus non-carriers. The Lancet. 2004; 364(9435): 703-705.

Whitt D, Salyers A. Bacterial pathogenesis: A molecular approach USA: American Sociey for Microbiology Press; 2002.

Wieser A, Schneider L, Jung J, Schubert S. MALDI-TOF MS in microbiological diagnostics-identification of microorganisms and beyond (mini review). Applied Microbiology and Biotechnology. 2012; 93(3): 965-974.

Wilke M, Lovering A, Strynadka C. β-lactam antibiotic resistance: a current structural perspective. Current Opinion in Microbiology. 2005; 8(5): 525-533.

Williams R. Healthy carriage of *Staphylococcus aureus*: its prevalence and importance. Bacteriology reviews. 1963; 27(96): 56-71.

Wilson J, Guy R, Elgohari S, Sheridan E, Davies J, Lamagni T, Pearson A. Trends in sources of meticillin-resistant *Staphylococcus aureus* (MRSA) bacteraemia: data from the national mandatory surveillance of MRSA bacteraemia in England, 2006-2009. Journal of Hospital Infection. 2011; 79(3): 211-217.

Wirtz C, Witte W, Wolz C, Goerke C. Insertion of host DNA into PVL-encoding phages of the *Staphylococcus aureus* lineage ST80 by intra-chromosomal recombination. Virology. 2010; 406(2): 322-327.

Witney A, Marsden G, Holden M, Stabler R, Husain S, Vass J, Butcher P, Hinds J, Lindsay J. Design, validation, and application of a seven-strain *Staphylococcus aureus* PCR product microarray for comparative genomics. Applied and Environmental Microbiology. 2005; 71(11): 7504-7514.

Woese C, Fox G, Zablen L, Uchida T, Bonen L, Pechman K, Lewis B, Stahl D. Conservation of primary structure in 16S ribosomal RNA. Nature. 1975; 254(5495): 83-86.

Wolters M, Rohde H, Maier T, Belmar-Campos C, Franke G, Scherpe S, Aepfelbacher M, Christner M. MALDI-TOF MS fingerprinting allows for discrimination of major methicillin-resistant *Staphylococcus aureus* lineages. International Journal of Medical Microbiology. 2011; 301(1): 64-68.

Wright A, Higginbottom A, Philippe D, Upadhyay A, Bagby S, Read R, Monk P, Partridge L. Characterisation of receptor binding by the chemotaxis inhibitory protein of *Staphylococcus aureus* and the effects of the host immune response. Molecular Immunology. 2007; 44(10): 2507-2517.

Wright G. Antibiotic resistance in the environment: a link to the clinic? Current Opinion in Microbiology. 2010; 13(5): 589-594.

Wright J. Mutation at VNTRs: Are minisatellites the evolutionary progeny of microsatellites? Genome. 1994; 37(2): 345-347.

Wyke A, Ward J, Hayes M. Synthesis of peptidoglycan *in vivo* in methicillin-resistant *Staphylococcus aureus*. European Journal of Biochemistry. 1982; 127(3): 553-558.

Wyllie D, Paul J, Crook D. Waves of trouble: MRSA strain dynamics and assessment of the impact of infection control. Journal of Antimicrobial Chemotherapy. 2011; 66(12): 2685-2688.

Yamaguchi T, Hayashi T, Takami H, Ohnishi M, Murata T, Nakayama K, Asakawa K, Ohara M, Komatsuzawa H, Sugai M. Complete nucleotide sequence of a *Staphylococcus aureus* exfoliative toxin B plasmid and identification of a novel ADP-ribosyltransferase, EDIN-C. Infection and Immunity. 2001; 69(12): 7760-7771.

Yarwood J, McCormick J, Paustian M, Orwin P, Kapur V, Schlievert P. Characterization and expression analysis of *Staphylococcus aureus* pathogenicity island 3. Implications for the evolution of staphylococcal pathogenicity islands. The Journal of Biological Chemistry. 2002; 277(15): 13138-13147.

Yoon H, Choi J, Kim C, Kim J, Song Y. A comparison of clinical features and mortality among methicillin-resistant and methicillin-sensitive strains of *Staphylococcus aureus* endocarditis. Yonsei Medical Journal. 2005; 46(4): 496-502.

Yoshida T, Kondo N, Hanifah Y, Hiramatsu K. Combined use of ribotyping, PFGE typing and IS*431* typing in the discrimination of nosocomial strains of methicillin-resistant *Staphylococcus aureus*. Microbiology and Immunology. 1997; 41(9): 687-695.

Zetola N, Francis J, Nuermberger E, Bishai W. Community-acquired meticillin-resistant *Staphylococcus aureus*: an emerging threat. The Lancet Infectious Diseases. 2005; 5: 275-286.

Zhang K, McClure J, Conly J. Enhanced multiplex PCR assay for typing of staphylococcal cassette chromosome *mec* types I to V in methicillin-resistant *Staphylococcus aureus*. Molecular and Cellular Probes. 2012; 26(5): 218-221.

Zhang K, McClure J, Elsayed S, Conly J. Novel staphylococcal cassette chromosome *mec* type, tentatively designated type VIII, harboring class A *mec* and type 4 *ccr* gene complexes in a canadian epidemic strain of methicillin-resistant *Staphylococcus aureus*. Antimicrobial Agents and Chemotherapy. 2009; 53(2): 531-540.

Zheng L, White R, Cash V, Jack R, Dean D. Cysteine desulfurase activity indicates a role for NIFS in metallocluster biosynthesis. Proceedings for the National Academy of Sciences of the United States of America. 1993; 90(7): 2754-2758.

Ziebuhr W, Krimmer V, Rachid S, Lößner I, Götz F, Hacker J. A novel mechanism of phase variation of virulence in *Staphylococcus* epidermidis: evidence for control of the polysaccharide intercellular adhesin synthesis by alternating insertion and excision of the insertion sequence element IS*256*. Molecular Microbiology. 1999; 32(2): 345-356.

Appendices

Appendix I. Characteristics of the 34 *S. aureus* isolates from the Glasgow collection at PHE, London

Isolate number	Bacteriophage type	PFGE type	Oxacillin	Mupirocin	*nuc* gene	*mecA* gene	*mupA* gene	Comments
DBHT 52	932/29/52/52A/80/95/77ih+	PF111a	R[*]	S[†]	P[‡]	P	N[§]	
DBHT 53	29/80/54/75/77/81+	PF121a	R	S	P	P	N	
DBHT 55	83C/75w/83Aw	PF016p	R	S	ND[‖]	ND	ND	EMRSA-16[¶]
DBHT 56	NT[#]	PF116a	B[**]	S	P	P	N	
DBHT 57	83Cih/932ih/29ih/52ih/52Aih/79ih/80ih/47ih/54ih/75ih/83A/84/85ih+	PF140a	R	S	ND	P	ND	
DBHT 58	ND	ND	ND	ND	ND	ND	ND	
DBHT 60	932w/29w/52w/3Cw/6w/88A/90w/94w	PF118a	B	HL[††]	P	P	P	
DBHT 61	75w	PF016b	R	S	ND	ND	ND	EMRSA-16
DBHT 62	83C/29ih/75w/77w/83Aw	PF016a	R	S	P	P	N	EMRSA-16
DBHT 63	29w/79w/54w/84w/81w	PF137a	S	S	P	N	N	
DBHT 64	ND	PF155b	R	S	ND	ND	ND	
DBHT 65	NT	PF132a	B	S	P	N	N	
DBHT 66	52ih/75w	PF016d	R	LL[‡‡]	ND	ND	ND	EMRSA-16
DBHT 67	29ih/75ih/77w/84ih/85ih	PF109a	R	S	ND	P	ND	
DBHT 68	NT	PF015b	R	S	ND	ND	ND	EMRSA-15
DBHT 69	6w	PF145a	R	S	P	P	N	
DBHT 70	83C/932w/52Aw/6/42Ew/47w/53w/54w/75w/83Aw/90w/81w	PF147a	R	S	ND	P	ND	
DBHT 71	932/29/52/52A/80/95/6ih/47ih/75ih/77/83Aih/84/88Aih/90ih+	PF134a	B	S	P	P	N	

Appendix I continued. Characteristics of the 34 *S. aureus* isolates from the Glasgow collection at PHE, London

Isolate number	Bacteriophage type	PFGE type	Oxacillin	Mupirocin	*nuc* gene	*mecA* gene	*mupA* gene	Comments
DBHT 72	932w/52w/52Aw/79w/80w/95w/6w/42Ew/47w/54w/83Aw/84w/90w/81w	PF135a	B	S	P	P	N	
DBHT 73	52Aw/3Aw/3Cw/55/71	PF127a	S	S	P	N	N	
DBHT 74	932ih/6ih/47ih/75ih/84ih/88Aih	PF001b	R	S	ND	P	ND	
DBHT 76	932/47ih/75ih/84ih/85ih+	PF141a	R	S	ND	P	ND	
DBHT 77	ND	PF153a	R	S	P	P	N	
DBHT 78	NT	PF015d	R	S	ND	ND	ND	EMRSA-15
DBHT 80	42Eih/47w/54w/75w/81w	PF126a	ND	HL	P	P	P	
DBHT 81	83C/932/52A/79/42Eih/47/53/54/75/77/84/85/88A/90/81/94+	PF098a	R	S	ND	P	ND	
DBHT 82	75w	PF015a	R	S	P	P	N	EMRSA-15
DBHT 83	83C/6/42E/54/75/77/83A/81+	PF105a	B	S	ND	P	ND	
DBHT 85	932/6/47/54ih/75/85/90+	PF117a	R	S	P	P	N	
DBHT 86	NT	PF107a	R	S	P	N	P	
DBHT 87	ND	PF151a	B	S	P	P	N	
DBHT 88	ND	PF149a	R	S	P	P	N	
DBHT 89	ND	PF157a	R	S	ND	P	ND	
DBHT 90	ND	PF156a	R	S	ND	P	ND	

Bacteriophage typing, PFGE, oxacillin and mupirocin minimum inhibitory concentrations, *nuc*, *mecA* and *mupA* PCRs were performed on a subset of the isolates by Staphylococcal Reference Laboratory (PHE, London). Oxacillin susceptibility tested by disk diffusion and Etest. PFGE, number indicates PFGE type and letter in lower case or number after indicates subtype. Subtypes differ by a few bands and are part of the same clone.

*R, Resistant.

†S, susceptible.

‡P, Positive.

§N, Negative.

||ND, No data.

¶EMRSA, Epidemic MRSA.

#NT, non-typeable.

**B, Borderline.

††HL, High level.

‡‡LL, Low level.

Appendix II. Two hundred and two *S. aureus* isolates obtained from the 2009 British Society for Antimicrobial Chemotherapy bacteraemia resistance surveillance programme

Isolate number	Ciprofloxacin	Erythromycin	Oxacillin	Piperacillin-Tazobactam	Teicoplanin	Vancomycin	*mecA* gene	*mupA* gene	Centre	Comments
DBHT 91	128	>128	128	128	1	1	P*	N†	32	HA‡
DBHT 92	64	0.25	>128	128	1	1	P	N	32	HA
DBHT 93	128	>128	64	64	1	1	P	N	32	HA
DBHT 94	128	>128	64	128	1	1	P	N	32	HA
DBHT 95	64	>128	>128	128	1	1	P	N	32	HA
DBHT 96	128	>128	64	64	1	1	P	N	32	HA
DBHT 97	64	>128	8	4	0.5	1	P	N	32	CA§
DBHT 98	128	>128	>128	>128	2	1	P	P	2	HA
DBHT 99	>128	0.5	>128	>128	4	2	P	N	2	HA
DBHT 100	32	>128	>128	64	1	2	P	N	3	HA
DBHT 101	64	>128	>128	128	1	1	P	N	3	CA
DBHT 102	128	>128	128	128	1	1	P	N	3	CA
DBHT 103	>128	>128	>128	128	2	1	P	N	3	HA
DBHT 104	128	>128	>128	128	1	1	P	N	3	HA
DBHT 105	64	>128	128	128	1	1	P	N	3	HA
DBHT 106	128	>128	128	128	1	2	P	N	4	HA
DBHT 107	>128	0.5	128	128	1	1	P	N	4	CA
DBHT 108	>128	>128	>128	>128	1	2	P	P	4	HA
DBHT 109	>128	>128	>128	128	1	2	P	N	4	HA
DBHT 110	128	0.5	64	64	1	1	P	N	5	HA
DBHT 111	>128	0.25	>128	32	1	1	P	N	12	HA

Appendix II continued. Two hundred and two *S. aureus* isolates obtained from the 2009 British Society for Antimicrobial Chemotherapy bacteraemia resistance surveillance programme

Isolate number	Ciprofloxacin	Erythromycin	Oxacillin	Piperacillin-Tazobactam	Teicoplanin	Vancomycin	*mecA* gene	*mupA* gene	Centre	Comments
DBHT 112	64	>128	>128	>128	1	2	P	N	12	HA
DBHT 113	16	0.5	128	128	1	1	P	N	8	CA
DBHT 114	16	0.25	128	64	1	1	P	N	8	CA
DBHT 115	128	>128	>128	128	1	1	P	N	8	CA
DBHT 116	64	>128	>128	128	0.5	<=0.5	P	N	8	CA
DBHT 117	16	0.5	>128	128	0.5	1	P	N	8	CA
DBHT 118	64	>128	128	128	0.5	1	P	N	8	ND[11]
DBHT 119	128	0.25	128	128	1	1	P	N	8	CA
DBHT 120	64	>128	128	64	1	1	P	N	10	CA
DBHT 121	128	>128	128	1	2	2	P	N	10	CA
DBHT 122	128	>128	128	64	1	1	P	N	10	CA
DBHT 123	128	>128	>128	>128	1	1	P	N	33	CA
DBHT 124	16	>128	>128	>128	1	1	P	N	33	HA
DBHT 125	64	>128	128	128	1	1	P	N	33	HA
DBHT 126	128	>128	>128	>128	1	2	P	N	33	HA
DBHT 127	128	>128	>128	128	4	2	P	N	33	HA
DBHT 128	>128	0.5	128	64	0.5	1	P	N	13	HA
DBHT 129	128	0.5	>128	>128	1	2	P	N	13	HA
DBHT 130	16	>128	>128	128	1	1	P	N	13	CA
DBHT 131	64	>128	>128	128	1	1	P	N	34	CA
DBHT 132	128	>128	>128	128	2	2	P	N	34	HA
DBHT 133	64	>128	>128	128	1	2	P	N	34	CA

Appendix II continued. Two hundred and two *S. aureus* isolates obtained from the 2009 British Society for Antimicrobial Chemotherapy bacteraemia resistance surveillance programme

Isolate number	Ciprofloxacin	Erythromycin	Oxacillin	Piperacillin-Tazobactam	Teicoplanin	Vancomycin	*mecA* gene	*mupA* gene	Centre	Comments
DBHT 134	128	>128	>128	128	2	1	P	N	34	HA
DBHT 135	64	>128	>128	128	1	1	P	N	34	HA
DBHT 136	16	0.5	64	64	1	2	P	P	34	HA
DBHT 137	128	>128	>128	128	1	1	P	N	30	HA
DBHT 138	>128	>128	>128	128	1	1	P	N	30	CA
DBHT 139	>128	>128	64	64	1	1	P	N	15	HA
DBHT 140	128	>128	>128	128	1	1	P	N	15	CA
DBHT 141	128	>128	128	128	1	1	P	N	15	HA
DBHT 142	128	>128	128	128	1	1	P	N	15	HA
DBHT 143	64	>128	128	128	1	1	P	N	15	HA
DBHT 144	0.5	0.5	16	2	1	2	P	N	15	HA
DBHT 145	128	0.25	128	32	1	1	P	N	15	CA
DBHT 146	>128	0.5	128	128	1	1	P	N	15	HA
DBHT 147	16	0.5	128	128	1	1	P	N	15	HA
DBHT 148	0.5	0.5	32	16	1	1	P	N	20	HA
DBHT 149	32	>128	>128	128	1	1	P	N	20	HA
DBHT 150	128	0.5	>128	128	1	1	P	N	20	HA
DBHT 151	16	0.25	128	64	1	1	P	N	20	CA
DBHT 152	128	>128	>128	128	1	1	P	N	21	CA
DBHT 153	64	>128	>128	128	1	1	P	N	22	HA
DBHT 154	64	>128	128	64	1	1	P	N	22	CA
DBHT 155	128	0.5	>128	>128	1	1	P	N	7	CA

Appendix II continued. Two hundred and two *S. aureus* isolates obtained from the 2009 British Society for Antimicrobial Chemotherapy bacteraemia resistance surveillance programme

Isolate number	Ciprofloxacin	Erythromycin	Oxacillin	Piperacillin-Tazobactam	Teicoplanin	Vancomycin	*mecA* gene	*mupA* gene	Centre	Comments
DBHT 156	128	0.25	8	4	1	1	P	N	5	CA
DBHT 157	128	>128	>128	128	1	1	P	N	26	CA
DBHT 158	>128	0.25	>128	128	0.5	1	P	N	26	CA
DBHT 159	128	0.25	128	128	1	1	P	N	7	HA
DBHT 160	128	0.5	>128	128	1	1	P	N	5	HA
DBHT 161	>128	>128	128	32	0.5	1	P	P	26	CA
DBHT 162	>128	>128	>128	128	1	1	P	N	7	CA
DBHT 163	0.5	0.5	64	8	2	2	P	N	5	HA
DBHT 164	128	>128	128	64	1	1	P	N	26	HA
DBHT 165	128	0.25	>128	128	1	1	P	N	7	CA
DBHT 166	128	0.5	128	128	1	1	P	N	26	HA
DBHT 167	128	0.5	>128	128	1	1	P	N	26	HA
DBHT 168	32	>128	128	64	1	1	P	N	10	CA
DBHT 169	>128	>128	>128	128	1	1	P	N	11	HA
DBHT 170	64	0.25	>128	64	1	1	P	N	11	HA
DBHT 171	128	0.5	>128	128	1	1	P	N	11	HA
DBHT 172	128	>128	>128	128	1	1	P	N	11	CA
DBHT 173	1	0.5	32	8	1	2	P	N	11	HA
DBHT 174	128	0.25	>128	>128	1	1	P	N	12	HA
DBHT 175	128	>128	128	64	0.5	<=0.5	P	N	11	HA
DBHT 176	128	>128	4	4	1	1	P	N	12	CA
DBHT 177	128	>128	128	128	1	1	P	N	12	HA

Appendix II continued. Two hundred and two *S. aureus* isolates obtained from the 2009 British Society for Antimicrobial Chemotherapy bacteraemia resistance surveillance programme

Isolate number	Ciprofloxacin	Erythromycin	Oxacillin	Piperacillin-Tazobactam	Teicoplanin	Vancomycin	*mecA* gene	*mupA* gene	Centre	Comments
DBHT 178	128	>128	128	64	0.5	1	P	N	12	CA
DBHT 179	64	0.5	64	32	0.5	1	P	N	22	CA
DBHT 180	128	>128	>128	128	2	1	P	N	22	HA
DBHT 181	128	0.25	>128	>128	1	1	P	N	23	CA
DBHT 182	128	>128	>128	128	0.5	1	P	N	28	HA
DBHT 183	16	>128	>128	128	1	1	P	N	28	CA
DBHT 184	16	>128	>128	128	0.5	1	P	N	28	HA
DBHT 185	128	>128	128	64	0.5	1	P	N	28	HA
DBHT 186	128	>128	>128	64	4	2	P	N	28	CA
DBHT 187	128	>128	128	32	1	1	P	N	28	HA
DBHT 188	64	>128	64	64	1	1	P	N	24	CA
DBHT 189	64	>128	>128	128	1	1	P	N	35	HA
DBHT 190	64	0.5	128	128	1	1	P	N	35	HA
DBHT 191	128	>128	>128	128	1	2	P	N	35	HA
DBHT 192	128	0.5	128	128	1	2	P	N	29	HA
DBHT 193	64	0.5	128	128	1	2	P	N	29	HA
DBHT 194	0.5	0.25	0.25	1	1	1	N	N	32	HA
DBHT 195	0.5	0.5	0.25	2	2	2	N	N	32	HA
DBHT 196	0.25	0.5	0.25	4	1	2	N	N	32	CA
DBHT 197	1	>128	1	4	1	2	N	N	32	HA
DBHT 198	0.5	0.25	0.25	2	1	1	N	N	2	HA
DBHT 199	64	0.25	0.5	4	1	1	N	N	2	CA

Appendix II continued. Two hundred and two *S. aureus* isolates obtained from the 2009 British Society for Antimicrobial Chemotherapy bacteraemia resistance surveillance programme

Isolate number	Ciprofloxacin	Erythromycin	Oxacillin	Piperacillin-Tazobactam	Teicoplanin	Vancomycin	*mecA* gene	*mupA* gene	Centre	Comments
DBHT 200	1	0.5	0.125	0.5	1	2	N	N	2	HA
DBHT 201	0.5	0.5	0.25	2	2	2	N	N	2	CA
DBHT 202	0.5	0.5	0.25	2	2	2	N	N	2	CA
DBHT 203	0.5	0.25	0.125	1	1	1	N	N	3	CA
DBHT 204	0.5	64	0.5	2	2	2	N	N	3	CA
DBHT 205	0.25	0.25	0.25	2	1	2	N	N	3	HA
DBHT 206	1	0.5	0.125	0.5	2	2	N	N	3	CA
DBHT 207	1	0.5	0.25	4	2	2	N	N	4	HA
DBHT 208	0.5	0.5	0.25	1	2	2	N	N	4	HA
DBHT 209	0.5	0.5	0.5	1	1	2	N	N	4	HA
DBHT 210	1	0.5	0.25	2	1	2	N	N	4	HA
DBHT 211	0.5	0.5	0.25	1	1	1	N	N	5	CA
DBHT 212	1	64	0.5	4	1	1	N	N	5	HA
DBHT 213	64	>128	0.25	2	1	1	N	N	5	CA
DBHT 214	0.25	0.25	0.25	2	0.5	1	N	N	5	CA
DBHT 215	0.25	8	0.5	2	1	2	N	N	6	HA
DBHT 216	0.25	0.5	0.25	0.5	1	2	N	N	6	CA
DBHT 217	1	0.5	0.25	2	1	2	N	N	6	CA
DBHT 218	0.5	0.5	0.25	4	1	2	N	N	6	CA
DBHT 219	0.5	0.5	0.25	2	1	1	N	N	7	HA
DBHT 220	0.5	0.25	0.25	2	1	1	N	N	7	CA
DBHT 221	0.5	0.5	0.25	2	2	2	N	N	7	CA

Appendix II continued. Two hundred and two *S. aureus* isolates obtained from the 2009 British Society for Antimicrobial Chemotherapy bacteraemia resistance surveillance programme

Isolate number	Ciprofloxacin	Erythromycin	Oxacillin	Piperacillin-Tazobactam	Teicoplanin	Vancomycin	*mecA* gene	*mupA* gene	Centre	Comments
DBHT 222	1	0.5	0.25	1	1	1	N	N	7	CA
DBHT 223	0.25	0.5	0.25	2	1	1	N	N	26	CA
DBHT 224	2	0.5	0.5	1	0.5	1	N	N	26	HA
DBHT 225	1	0.25	0.5	2	1	1	N	N	26	CA
DBHT 226	0.5	0.25	0.25	1	1	1	N	N	26	HA
DBHT 227	1	0.5	0.25	1	1	2	N	N	8	CA
DBHT 228	1	0.25	0.5	2	2	2	N	N	8	CA
DBHT 229	0.5	0.5	0.25	0.5	1	1	N	N	8	HA
DBHT 230	1	0.25	0.25	1	1	1	N	N	8	CA
DBHT 231	0.5	0.5	0.25	2	1	2	N	N	10	CA
DBHT 232	1	>128	0.5	2	1	1	N	N	10	CA
DBHT 233	1	0.5	0.25	0.5	1	1	N	N	10	CA
DBHT 234	0.25	0.5	0.125	1	1	2	N	N	10	CA
DBHT 235	128	0.25	0.25	1	1	2	N	N	11	CA
DBHT 236	0.25	0.5	0.25	1	1	1	N	N	11	CA
DBHT 237	1	0.5	0.25	0.5	1	1	N	N	11	HA
DBHT 238	1	0.5	0.25	2	1	2	N	N	11	HA
DBHT 239	1	0.25	0.25	1	1	2	N	N	12	CA
DBHT 240	0.5	0.5	0.5	2	1	2	N	N	12	HA
DBHT 241	0.25	0.5	2	2	1	1	N	N	12	HA
DBHT 242	1	0.25	0.25	1	1	1	N	N	12	CA
DBHT 243	4	0.25	0.25	1	2	2	N	N	33	HA

Appendix II continued. Two hundred and two *S. aureus* isolates obtained from the 2009 British Society for Antimicrobial Chemotherapy bacteraemia resistance surveillance programme

Isolate number	Ciprofloxacin	Erythromycin	Oxacillin	Piperacillin-Tazobactam	Teicoplanin	Vancomycin	*mecA* gene	*mupA* gene	Centre	Comments
DBHT 244	1	0.5	0.5	1	2	2	N	N	33	CA
DBHT 245	1	0.5	0.5	1	1	2	N	N	33	CA
DBHT 246	0.5	0.5	0.25	1	2	2	N	N	33	CA
DBHT 247	0.5	0.5	0.5	2	1	1	N	N	13	CA
DBHT 248	1	0.5	0.25	0.5	1	2	N	N	13	CA
DBHT 249	0.5	0.5	0.25	0.5	1	2	N	N	13	CA
DBHT 250	0.5	8	0.25	1	1	1	N	N	13	HA
DBHT 251	4	0.5	0.25	0.5	2	2	N	N	34	CA
DBHT 252	0.25	0.5	0.5	1	2	2	N	N	34	CA
DBHT 253	0.5	0.5	0.25	1	2	2	N	N	34	HA
DBHT 254	0.5	0.5	0.5	1	1	2	N	N	34	HA
DBHT 255	0.5	0.5	0.25	2	1	2	N	N	30	CA
DBHT 256	0.5	0.5	0.25	1	1	1	N	N	30	HA
DBHT 257	1	0.5	0.25	1	1	2	N	N	30	CA
DBHT 258	0.5	0.5	0.25	1	0.5	1	N	N	30	CA
DBHT 259	0.125	0.25	0.25	1	2	1	N	N	15	CA
DBHT 260	0.25	0.5	0.25	2	1	2	N	N	15	HA
DBHT 261	0.5	0.5	0.125	1	1	1	N	N	15	HA
DBHT 262	1	0.5	0.5	4	1	2	N	N	15	HA
DBHT 263	0.25	0.5	0.125	0.5	2	2	N	N	20	HA
DBHT 264	0.5	0.5	0.25	1	2	1	N	N	20	HA
DBHT 265	0.5	0.5	0.25	2	0.5	1	N	N	20	CA

Appendix II continued. Two hundred and two *S. aureus* isolates obtained from the 2009 British Society for Antimicrobial Chemotherapy bacteraemia resistance surveillance programme

Isolate number	Ciprofloxacin	Erythromycin	Oxacillin	Piperacillin-Tazobactam	Teicoplanin	Vancomycin	*mecA* gene	*mupA* gene	Centre	Comments
DBHT 266	0.5	0.5	0.5	2	2	2	N	N	20	HA
DBHT 267	1	0.5	0.25	4	1	2	N	N	21	HA
DBHT 268	0.5	4	0.5	2	1	2	N	N	21	HA
DBHT 269	1	2	0.25	2	1	2	N	N	21	CA
DBHT 270	8	32	0.5	2	1	1	N	N	21	HA
DBHT 271	1	>128	0.5	4	1	1	N	N	22	HA
DBHT 272	0.5	0.5	0.25	0.5	1	1	N	N	22	CA
DBHT 273	0.5	0.5	0.5	2	2	2	N	N	22	HA
DBHT 274	0.5	0.5	0.25	2	1	1	N	N	22	CA
DBHT 275	0.5	0.5	0.25	1	1	2	N	N	22	CA
DBHT 276	1	0.5	0.125	1	1	2	N	N	23	HA
DBHT 277	0.5	0.5	0.25	4	1	2	N	N	23	HA
DBHT 278	1	0.5	0.25	0.5	1	2	N	N	23	CA
DBHT 279	0.5	0.5	0.5	4	2	2	N	N	23	CA
DBHT 280	1	0.5	0.25	2	2	2	N	N	28	HA
DBHT 281	0.5	0.5	0.5	2	1	2	N	N	28	CA
DBHT 282	0.25	0.5	0.25	0.5	1	1	N	N	28	HA
DBHT 283	0.25	0.5	0.125	0.5	1	2	N	N	28	CA
DBHT 284	0.5	0.5	0.5	2	1	2	N	N	24	CA
DBHT 285	1	0.5	0.25	4	2	1	N	N	24	HA
DBHT 286	0.5	0.5	0.25	2	1	2	N	N	24	HA
DBHT 287	0.25	0.5	0.125	0.25	1	2	N	N	24	CA

Appendix II continued. Two hundred and two *S. aureus* isolates obtained from the 2009 British Society for Antimicrobial Chemotherapy bacteraemia resistance surveillance programme

Isolate number	Ciprofloxacin	Erythromycin	Oxacillin	Piperacillin-Tazobactam	Teicoplanin	Vancomycin	*mecA* gene	*mupA* gene	Centre	Comments
DBHT 288	1	0.5	0.125	0.5	1	2	N	N	35	HA
DBHT 289	64	>128	0.25	2	1	1	N	N	35	HA
DBHT 290	128	>128	0.25	2	1	1	N	N	35	HA
DBHT 291	1	0.5	0.5	1	1	2	N	N	35	HA
DBHT 292	1	0.5	0.125	1	1	2	N	N	29	HA

Antimicrobial susceptibilities to six antimicrobials and *mecA* and *mupA* PCRs were performed on the isolates by Antimicrobial Resistance and Healthcare Associated Infections Reference Unit, PHE, London.

*P, positive.

†N, negative.

‡HA, hospital-associated.

§CA, community-associated.

||ND, no data.

Appendix III. Antimicrobial susceptibilities to eight antimicrobials for six *S. aureus* isolates obtained from the Clinical Microbiology and Public Health Laboratory

Isolate number	Oxacillin	Ciprofloxacin	Erythromycin	Fusidic acid	Gentamicin	Mupirocin	Neomycin	Rifampicin
DBHT 293	R*	R	R	S†	R	R	S	S
DBHT 294	R	R	R	S	R	R	S	S
DBHT 296	R	R	R	S	R	R	R	S
DBHT 300	R	R	R	S	R	R	S	S
DBHT 301	R	R	R	S	R	R	R	S
DBHT 302	R	R	R	S	R	R	S	S

*R, Resistant.

†S, Susceptible.

Appendix IV. Sequences of PCR primers used in this study

PCR	Primer name	Primer sequence (5' to 3')	Reference
16S rRNA gene PCR	ANTIF	AGAGTTTGATCCTGGCTCAG	(Edwards *et al.*, 1989)
	1392R	ACGGGCGGTGTGTACAAG	(Lane *et al.*, 1985) (Lane, 1991)
16S rRNA gene sequence determination	357F	CTCCTACGGGAGGCAGCAG	(Lane, 1991)
	3R	GTTGCGCTCGTTGCGGGACT	(Coenye *et al.*, 1999)
arcC	*arcC*-F	TTGATTCACCAGCGCGTATTGTC	(Enright *et al.*, 2000)
	arcC-R	AGGTATCTGCTTCAATCAGCG	
aroE	*aroE*-F	ATCGGAAATCCTATTTCACATTC	
	aroE-R	GGTGTTGTATTAATAACGATATC	
glpF	*glpF*-F	CTAGGAACTGCAATCTTAATCC	
	glpF-R	TGGTAAAATCGCATGTCCAATTC	
gmk	*gmk*-F	ATCGTTTTATCGGGACCATC	
	gmk-R	TCATTAACTACAACGTAATCGTA	
pta	*pta*-F	GTTAAAATCGTATTACCTGAAGG	
	pta-R	GACCCTTTTGTTGAAAAGCTTAA	
tpi	*tpi*-F	TCGTTCATTCTGAACGTCGTGAA	
	tpi-R	TTTGCACCTTCTAACAATTGTAC	
yqiL	*yqiL*-F	CAGCATACAGGACACCTATTGGC	
	yqiL-R	CGTTGAGGAATCGATACTGGAAC	
SCC*mec*-*orfX* junction	SCC*mec*-*orf*XF	CATGAAAATCACCATTTTAGCTG	This study
	SCC*mec*-*orf*XR	ATAGTTCCCAGTAGCAACCT	This study
SCC*mec*-*orfX* junction nested	SCC*mec*-*orf*X2F	TCCAGACGAAAAAGCACCAGAAAA	This study
	SCC*mec*-*orf*X2R	TTGAAATGAAAGACTGCGGAGGCT	This study
	SCC*mec*-*orf*X3R	ACATTATTGGTACGTTTCTCTTTA	This study
orfX	*orfX*F	CGCATAATCTTAAATGCCT	This study
	*orfX*R	ATGAAAATCACCATTTTAGC	This study
SCC*mec* type VI	*IS1272J*-F	GAAGCTTTGGGCGATAAAGA	(Chen *et al.*, 2009)
	IS1272J-R	GCACTGTCTCGTTTAGACCAATC	
	ccrB4-F	CGAAGTATAGACACTGGAGCGATA	
	ccrB4-R	GCGACTCTCTTGGCGTTTA	

Appendix IV continued. Sequences of PCR primers used in this study

PCR	Primer name	Primer sequence (5' to 3')	Reference
SCC*mec* type VII	*ccrC2-F2*	ATAAGTTAAAAGCACGACTCA	(Higuchi *et al.*, 2008)
	ccrC2-R2	TTCAATCCTATTTTTCTTTGTG	
	ccrC8-F	GCATGGGTACTCAATCCA	
	ccrC8-R	GGTTGTAATGGCTTTGAGG	
	mecC2-F	ATCAGTTCATTGCTCACGATATGTGTA	
	mecC2-R	CAATACGCCATTTGTAATAAGCTTTT	
SCC*mec* type VIII	*mecI-F*	CCCTTTTTATACAATCTCGTT	(McClure *et al.*, 2010)
	mecI-R	ATATCATCTGCAGAATGGG	
	ccr4-Fd	ATCGCTCATTATGGATACYGC	
	ccr4-R2	CAAAACAACCTTTTCTATAACG	
	VIII-F3	CAATATTGATTTCCTTCATCGTTTACCTCC	
	VIII-R3	GAGCATCATAAGAAGCAATTTTATGTTACGC	
SCC*mec* non-typeable	*mA1*	TGCTATCCACCCTCAAACAGG	(Kondo *et al.*, 2007)
	mA2	AACGTTGTAACCACCCCAAGA	
	α1	AACCTATATCATCAATCAGTACGT	
	α2	TAAAGGCATCAATGCACAAACACT	
	α3	AGCTCAAAAGCAAGCAATAGAAT	
	βc	ATTGCCTTGATAATAGCCITCT	
	α4.2	GTATCAATGCACCAGAACTT	
	β4.2	TTGCGACTCTCTTGGCGTTT	
	γR	CCTTTATAGACTGGATTATTCAAAATAT	
	γF	CGTCTATTACAAGATGTTAAGGATAAT	
	mI6	CATAACTTCCCATTCTGCAGATG	
	IS7	ATGCTTAATGATAGCATCCGAATG	
	IS2(iS-2)	TGAGGTTATTCAGATATTTCGATGT	
	mA7	ATATACCAAACCCGACAACTACA	

Forward and reverse primer sequences used in PCR and sequence determination assays in this study.

Appendix V. Sequences obtained from the NCBI database used in this study

Accession no.	Strain name	Group
NC_002745	N315	A, D, E, F, G
NC_009782	Mu3	A, D, F, G
NC_002758	Mu50	A, D, E, F, G
NC_007622	RF122	A, D, E, F, G
NC_010079	USA300_TCH1516	A, D, F, G
NC_003923	MW2	A, D, E, F, G
NC_013450	ED98	A, D, F, G
NC_009641	Newman	A, D, F, G
NC_007793	USA300_FPR3757	A, D, E, F, G
NC_002951	COL	A, D, E, F, G
NC_002953	MSSA476	A, D, E, F, G
NC_002952	MRSA252	A, D, E, F, G
NC_009632	JH1	A, D, F, G
NC_009487	JH9	A, D, E, F, G
NC_007795	NCTC 8325	A, D, E, F, G
AB425824	JCSC6670	B
CP000255	USA300_FPR3757	B
AB425823	JCSC6668	B
BA000033	MW2	B
CP000730	USA300_TCH1516	B
AB121219	JCSC3624	C
AB033763	NCTC 10442	C
D86934	N315	C
AB425427	V14	C
AB063172	CA05	C
AB063173	JCSC1978	C
NC_017331	TW20	D, G
NC_017338	JKD6159	D, G
CP003166	M013	D, G
HE681097	HO50960412	D, G
NC_017349	LGA251	G
NC_017333	ST398	G

Group A, used to design *gmk* PCR primers; Group B, used to design SCC*mec*-*orfX* PCR primers; Group C, used to design *orfX* PCR primers; Group D, used to calculate mean size of *S. aureus* genome; Group E, used to generate *in silico* FAFLP profiles; Group F, used to search for presence of insertion sequences; Group G, used to generate *in silico* optical maps.

Appendix VI. Results of *spa* typing

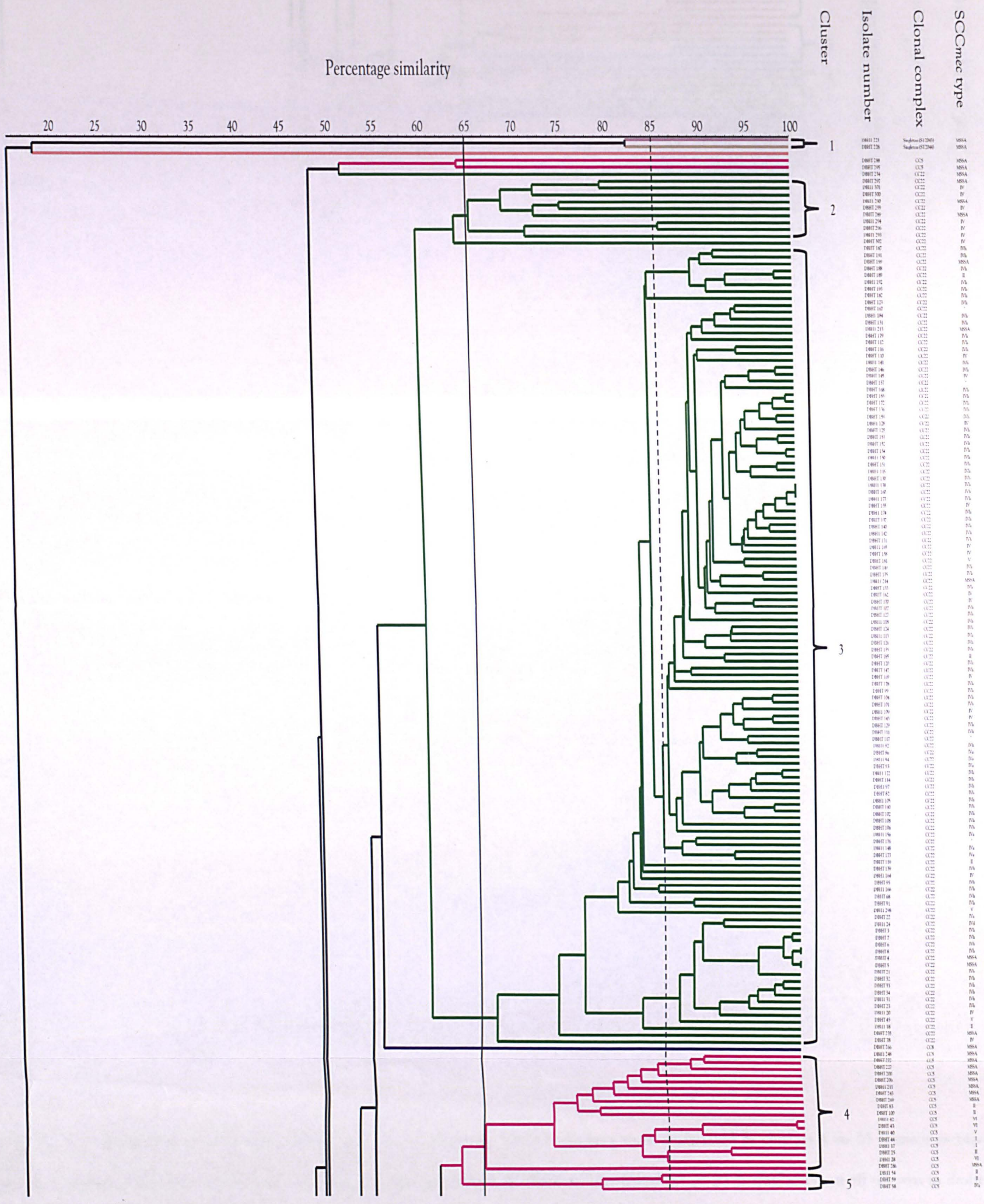
Isolate number	*spa* type	*spa* repeats
DBHT 1	t315	26-22-17-20-17-12-17-17-16
DBHT 2	t315	26-22-17-20-17-12-17-17-16
DBHT 3	t8530	26-23-23-13-23-31-29-16-31-29-17-25-17-25-16-28
DBHT 6	t8530	26-23-23-13-23-31-29-16-31-29-17-25-17-25-16-28
DBHT 7	t8530	26-23-23-13-23-31-29-16-31-29-17-25-17-25-16-28
DBHT 8	t8530	26-23-23-13-23-31-29-16-31-29-17-25-17-25-16-28
DBHT 12	t131	07-23-12-34-33-34
DBHT 13	t186	07-12-21-17-13-13-34-34-33-34
DBHT 14	t202	11-17-23-17-17-16-16-25
DBHT 15	t190	11-17-34-24-34-22-25
DBHT 16	t190	11-17-34-24-34-22-25
DBHT 17	t001	26-30-17-34-17-20-17-12-17-16
DBHT 18	t065	09-02-16-34-13-17-34-16-34
DBHT 19	t128	07-23-23-21-16-34-33-13
DBHT 20	t022	26-23-13-23-31-29-17-31-29-17-25-17-25-16-28
DBHT 21	t032	26-23-23-13-23-31-29-17-31-29-17-25-17-25-16-28
DBHT 22	t005	26-23-13-23-31-05-17-25-17-25-16-28
DBHT 23	t032	26-23-23-13-23-31-29-17-31-29-17-25-17-25-16-28
DBHT 24	t005	26-23-13-23-31-05-17-25-17-25-16-28
DBHT 25	t002	26-23-17-34-17-20-17-12-17-16
DBHT 26	t008	11-19-12-21-17-34-24-34-22-25
DBHT 27	t127	07-23-21-16-34-33-13
DBHT 28	t311	26-23-17-34-20-17-12-17-16
DBHT 29	t004	09-02-16-13-13-17-34-16-34
DBHT 30	t019	08-16-02-16-02-25-17-24
DBHT 31	t9238	26-23-31-31-29-17-31-29-17-25-17-25-16-28
DBHT 32	t9238	26-23-31-31-29-17-31-29-17-25-17-25-16-28
DBHT 33	t032	26-23-23-13-23-31-29-17-31-29-17-25-17-25-16-28
DBHT 34	t032	26-23-23-13-23-31-29-17-31-29-17-25-17-25-16-28
DBHT 39	t630	08-16-02-16-34-17-34-16-34
DBHT 40	t640	26-23-17-12-16
DBHT 41	t026	08-16-34
DBHT 42	t548	26-23-17-34-17-20-17-12-16
DBHT 43	t5490	26-23-17-34-17-20-17-12-17-274
DBHT 44	t026	08-16-34
DBHT 45	t474	26-23-13-23-31-05-17-25-16-28
DBHT 46	t127	07-23-21-16-34-33-13
DBHT 47	t267	07-23-12-21-17-34-34-34-33-34
DBHT 49	t316	04-20-17-31-16-34
DBHT 50	t316	04-20-17-31-16-34
DBHT 51	t018	15-12-16-02-16-02-25-17-24-24-24
DBHT 52	t026	08-16-34
DBHT 53	t091	07-23-21-17-34-12-23-02-12-23
DBHT 54	t002	26-23-17-34-17-20-17-12-17-16

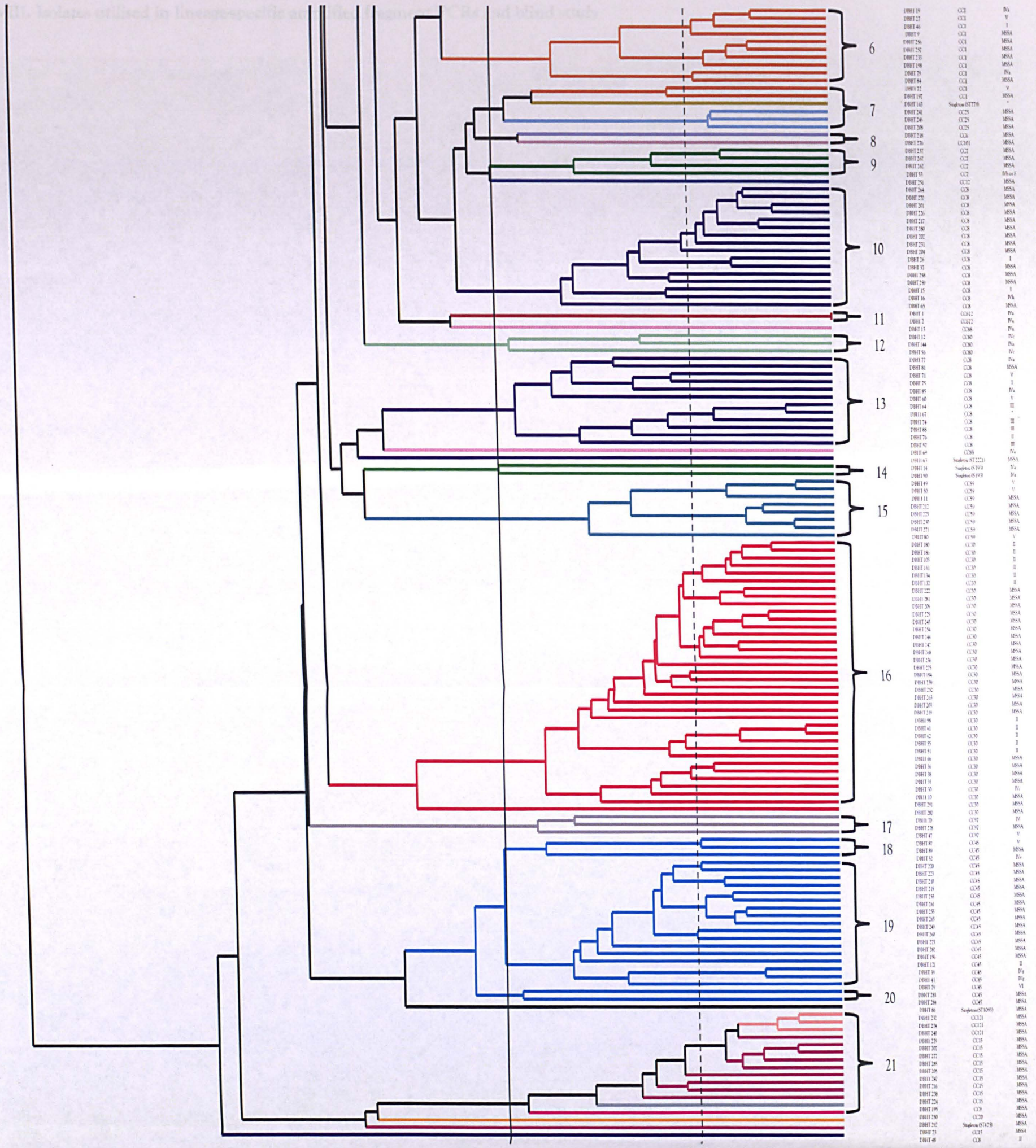
Appendix VI continued. Results of *spa* typing

Isolate number	*spa* type	*spa* repeats
DBHT 55	t9239	15-21-16-02-25-17-24-24-24
DBHT 56	t044	07-23-12-34-34-33-34
DBHT 57	t037	15-12-16-02-25-17-24
DBHT 58	t450	26-23-17-34-16
DBHT 59	t002	26-23-17-34-17-20-17-12-17-16
DBHT 60	t3209	04-34-24-34-22-25
DBHT 61	t018	15-12-16-02-16-02-25-17-24-24-24
DBHT 62	t018	15-12-16-02-16-02-25-17-24-24-24
DBHT 64	t037	15-12-16-02-25-17-24
DBHT 66	t018	15-12-16-02-16-02-25-17-24-24-24
DBHT 67	t037	15-12-16-02-25-17-24
DBHT 68	t032	26-23-23-13-23-31-29-17-31-29-17-25-17-25-16-28
DBHT 69	t2311	07-12-12-21-17-13-13-34-34-33-34
DBHT 70	t1236	26-23-12-21-17-34-34-34-33-34
DBHT 71	t190	11-17-34-24-34-22-25
DBHT 72	t5073	26-23-13-21-17-34-34-34-33-02-02-34
DBHT 74	t037	15-12-16-02-25-17-24
DBHT 75	t008	11-19-12-21-17-34-24-34-22-25
DBHT 76	t037	15-12-16-02-25-17-24
DBHT 77	t008	11-19-12-21-17-34-24-34-22-25
DBHT 78	t032	26-23-23-13-23-31-29-17-31-29-17-25-17-25-16-28
DBHT 79	t128	07-23-23-21-16-34-33-13
DBHT 80	t316	04-20-17-31-16-34
DBHT 82	t022	26-23-13-23-31-29-17-31-29-17-25-17-25-16-28
DBHT 83	t002	26-23-17-34-17-20-17-12-17-16
DBHT 85	t008	11-19-12-21-17-34-24-34-22-25
DBHT 87	t1081	08-16-02-43-34-17-34
DBHT 88	t138	08-16-02-25-17-24
DBHT 90	t202	11-17-23-17-17-16-16-25

***spa* type and *spa* repeats for 73 isolates. *spa* repeats represent the size and number of repeating units which are found in the polymorphic X region of the *spa* gene. The size and number of repeating units constitute a specific *spa* type.**

Appendix VII. Cluster analysis of UPGMA-derived dendrogram based on *Bgl*II+A and *Csp*6I+0 FAFLP data





The dendrogram was derived for all 302 study isolates and an *E. coli* strain. FAFLP clusters were designated as clusters 1 to 21 (shown as black numbers), based on 65 % similarity cut-off (shown as solid vertical black line). Sub-clusters were based on an 85 % similarity cut-off (shown as dashed vertical black line). The clonal complex and SCC*mec* type of isolates is shown on the right of the cluster number. Isolates assigned to the same clonal complex were highlighted in a single colour.

*Indicate SCC*mec* non-typeable.

Appendix VIII. Isolates utilised in lineage-specific amplified fragment PCRs and blind study

Isolates utilised in lineage-specific AF* PCRs	Isolates utilised in blind study	Clonal complex
DBHT 19, DBHT 79, DBHT 233	DBHT 19, DBHT 46, DBHT 233, DBHT 257	CC1
DBHT 25, DBHT 40, DBHT 59, DBHT 206, DBHT 272, DBHT 290	DBHT 28, DBHT 54, DBHT 59, DBHT 272, DBHT 288	CC5
DBHT 16, DBHT 67, DBHT 76, DBHT 88, DBHT 201, DBHT 259	DBHT 16, DBHT 48, DBHT 67, DBHT 76, DBHT 85, DBHT 88, DBHT 202, DBHT 231	CC8
DBHT 205, DBHT 247, DBHT 279	DBHT 205, DBHT 277, DBHT 279, DBHT 285	CC15
DBHT 18, DBHT 94, DBHT 130, DBHT 181, DBHT 235	DBHT 7, DBHT 91, DBHT 130, DBHT 140, DBHT 172, DBHT 199, DBHT 213, DBHT 235, DBHT 297	CC22
DBHT 30, DBHT 66, DBHT 161, DBHT 222, DBHT 239, DBHT 252, DBHT 287	DBHT 132, DBHT 161, DBHT 222, DBHT 239, DBHT 254, DBHT 281, DBHT 287	CC30
DBHT 87, DBHT 121, DBHT 255	DBHT 121, DBHT 210, DBHT 255, DBHT 273	CC45
DBHT 49, DBHT 212, DBHT 225	DBHT 11, DBHT 80, DBHT 212, DBHT 230	CC59
DBHT 47, DBHT 70, DBHT 278	DBHT 70, DBHT 278	CC97
DBHT 1	DBHT 2	CC672

*AF, amplified fragment

Appendix IX. Isolates utilised in Pyrosequencing™ assay

Isolate number	MLST
DBHT 72	CC1
DBHT 198*	CC1
DBHT 58*	CC5
DBHT 100*	CC5
DBHT 204	CC8
DBHT 226	CC8
DBHT 207	CC15
DBHT 22*	CC22
DBHT 31*	CC22
DBHT 45*	CC22
DBHT 96*	CC22
DBHT 118*	CC22
DBHT 138*	CC22
DBHT 173*	CC22
DBHT 214*	CC22
DBHT 302	CC22
DBHT 55*	CC30
DBHT 236*	CC30
DBHT 29	CC45
DBHT 196	CC45
DBHT 260*	CC45
DBHT 50*	CC59
DBHT 2	CC672

*Isolates assigned to a lineage based on lineage-specific SNPs and point mutations.

Isolates utilised in Pyrosequencing™ assay targeting 95 bp of lineage-specific 307 bp amplified fragment.

Appendix X. Loci amongst 15 whole-genome sequences which encode a transposase

Strain	No. of regions	Transposase regions
COL	7	36353-37027, 42958-44481, 467453-467725, 1436659-1436955, 1894859-1895344, 1932920-1934239, 1972991-1974310
ED98	24	53099-53590, 53556-54062, 399312-399584, 981105-982280, 982249-982425, 1093058-1094377, 1151848-1153167, 1408479-1408661, 1421603-1422922, 1713665-1714984, 1746451-1746771, 1746902-1747414, 1862395-1862559, 1862998-1863681, 1877982-1878551, 1879582-1880082, 1937411-1938763, 1976778-1977281, 1977349-1977798, 2140965-2142284, 2223290-2223460, 2223453-2223611, 2303723-2304049, 2738091-2739410
JH1	13	51535-51912, 64441-64794, 64779-65093, 718701-720020, 753283-754602, 972729-974048, 1263536-1264855, 1812284-1812661, 1963390-1964073, 2023726-2025045, 2038447-2039766, 2240636-2241955, 2478139-2479458
JH9	11	473451-473723, 718825-720144, 753407-754726, 972854-974173, 1263662-1264981, 1520180-1520476, 1980101-1980601, 2023851-2025170, 2038572-2039891, 2240761-2242080, 2478263-2479582
MRSA252	46	36403-37077, 41749-42423, 57680-58057, 58064-59956, 59953-61038, 70097-70939, 70924-71238, 89496-89795, 89837-90643, 91351-92298, 449607-449954, 733183-764882, 734448-735122, 742098-743744, 758443-760089, 763228-763605, 763700-764374, 973598-975244, 998073-999020, 1004249-1005037, 1006111-1007757, 1029169-1029300, 1182701-1184347, 1215677-1216624, 1351329-1352349, 1367414-1368361, 1409057-1409845, 1431432-1433078, 1493459-1494406, 1798303-1798680, 1798687-1800579, 1800576-1801661, 1909322-1910764, 1972462-1972606, 1993384-1993590,2119976-2121622, 2175406-2175627, 2245178-2245861, 2304037-2305684, 2306521-2307052, 2325364-2327010, 2392632-2393420, 2543734-2545380, 2690769-2691193, 2797873-2798820, 2890689-2891645
MSSA476	12	58948-59577, 59745-59855, 140437-140921, 1204084-1204269, 1426098-1426391, 1442115-1442596, 1459026-1459604, 1872598-1872807, 1872849-1873153, 2011686-2011897, 2205996-2206527, 2359810-2360145

Appendix X continued. Loci amongst 15 whole-genome sequences which encode a transposase

Mu3	31	36436-37110, 42007-42456, 57753-58130, 58137-60029, 60026-61111, 70170-70547, 70659-71012, 426640-427959, 460237-461427, 463952-464248, 703994-705313, 971634-972953, 1216813-1218132, 1252941-1253153, 1473328-1473624, 1765447-1765824, 1765831-1767723, 1767720-1768805, 1815037-1816356, 1841027-1842346, 1918797-1919147, 1919323-1919571, 1919490-1920269, 1934570-1935139, 1935161-1935415, 1936170-1936670, 1993998-1995317, 2071108-2071281, 2212484-2213803, 2506985-2508304, 2638860-2640179
Mu50	32	36436-37110, 41782-42456, 57753-58130, 58137-60029, 60026-61111, 70170-70547, 70659-71012, 70997-71311, 426637-427956, 460233-461423, 463948-464244, 702549-703868, 970190-971509, 1215412-1216731, 1251541-1251753, 1471928-1472224, 1764047-1764424, 1764431-1766323, 1766320-1767405, 1813637-1814956, 1839628-1840947, 1917398-1917748, 1917924-1918172, 1918091-1918870, 1933171-1933740, 1933762-1934016, 1934771-1935271, 1992601-1993920, 2069711-2069884, 2210846-2212165, 2505347-2506666, 2637221-2638540
MW2	12	36432-37106, 60241-60870, 61038-61148, 1175411-1175596, 1397643-1397837, 1413561-1413755, 1413773-1414165, 1430664-1431050, 1893263-1893472, 1893675-1893818, 2032439-2032612, 2380849-2381163
N315	30	36435-37109, 41781-42455, 57792-58169, 58176-60068, 60065-61150, 70209-70586, 70698-71051, 426680-427999, 438190-438486, 678300-679619, 867008-868093, 868090-869982, 869989-870366, 893906-895225, 1139084-1140403, 1175212-1175424, 1395599-1395895, 1687654-1688031, 1688038-1689930, 1689927-1691012,1761714-1763033, 1839605-1839955, 1840131-1840379, 1840298-1841077, 1855378-1855947, 1855969-1856223, 1914806-1916125, 1991916-1992089, 2134414-2135733, 2566795-2568114

Appendix X continued. Loci amongst 15 whole-genome sequences which encode a transposase

NCTC 8325	16	115867-116088, 264380-264784, 1111031-1111240, 1332886-1333182, 1348900-1349016, 1349111-1349503, 1705674-1706480, 1706498-1707325, 1813964-1814788, 1814821-1815429, 1820862-1821347, 1832051-1832260, 1832463-1832606, 1897329-1898531, 2237284-2238486, 2425635-2425991
Newman	20	115844-116065, 433128-433400, 1210988-1211197, 1432841-1433137, 1448855-1449049, 1449067-1449459, 1762474-1763280, 1763298-1764125, 1852428-1852919, 1870802-1871626, 1871659-1872267, 1877700-1878185, 1889301-1889444, 1954178-1955497, 2294897-2296135, 2322168-2322968, 2322965-2323285, 2323433-2323792, 2436373-2436687, 2483472-2483828
RF122	14	775800-776813, 812335-812604, 922627-923943, 979690-980364, 1268707-1269489, 1269528-1270028, 1421380-1422054, 1721534-1722850, 2020407-2020628, 2184629-2185453, 2185450-2185950, 2309331-2309831, 2309828-2310652, 2341147-2342463
USA300_FPR3757	11	35877-36551, 42442-43953, 67966-68790, 78432-79445, 318970-319374, 1443183-1443668, 1917170-1917571, 1993601-1994956, 2290382-2291620, 2431859-2432173, 2478538-2478894
USA300_TCH1516	11	35877-36551, 42442-43965, 67885-68736, 78378-79391, 318865-319269, 443428-443700, 1443035-1443427, 1457130-1457615, 1917920-1918321, 1994386-1995747, 2290812-2292014

*Number of regions amongst each of 15 whole-genome sequences (WGSs) which encoded a transposase. Regions were identified using Mauve software based on the search for the term 'transposase' amongst annotations available on the NCBI database (http://www.ncbi.nlm.nih.gov).

†Nucleotide positions of regions which encoded a transposase amongst the 15 WGSs.