University of Oxford

Staphylococcus aureus: The host-organism relationship

Ruth R. Miller
Linacre College

Nuffield Department of Medicine

A thesis submitted for the degree of Doctor of Philosophy

May 2011

Staphylococcus aureus: The host-organism relationship

Ruth Miller, Linacre College
DPhil Thesis, Trinity Term 2011

Abstract

Staphylococcus aureus is a worldwide leading cause of skin and soft tissue, bone and joint, and bloodstream infection. Despite this, *S. aureus* is also a harmless commensal in about one third of the population, although carriage is a risk factor for subsequent disease. *S. aureus* has evolved resistance to several antibiotics, including meticillin, resulting in meticillin-resistant *S. aureus* (MRSA), which in the UK largely consists of two epidemic lineages. In spite of much research, substantial aspects of the epidemiology and biology of *S. aureus* are still poorly understood.

In investigating the *S. aureus* host-organism relationship, this thesis has three aims. To explore the interface between community and hospital-acquired *S. aureus*; to investigate the carriage dynamics of *S. aureus* in the community; and to use population genetic methods to study epidemic hospital associated *S. aureus* lineages.

Case-control studies comparing hospital and community-acquired MRSA revealed that the majority of UK MRSA remains healthcare associated, with community-acquired MRSA reliably identified in only 0.2% of individuals. However, an additional 0.2% of individuals also carried "feral" MRSA with molecular characteristics identical to hospital-associated strains, but in hosts with no healthcare risk factors. To further investigate *S. aureus* carriage dynamics in the community, a carriage study was designed to collect detailed host factor information and correlate this with *S. aureus* carriage over time. In the study 32% of participants carried *S. aureus* of which the majority carried for over one year. Younger age was associated with transient carriage, including *S. aureus* acquisition in individuals who were initially negative. Finally, whole-genome sequencing of two epidemic *S. aureus* lineages indicated rapid clonal expansion of MRSA and clear geographic and temporal genetic structure. One particularly closely related cluster of strains may provide a genetic explanation for an MRSA outbreak in Brighton.

Acknowledgements

When I first started working on Staph I had never picked up a pipette or even heard of Python, so to get this far I have required a lot of help, and therefore have many people to thank. First and foremost, thank you to my supervisors: Rory Bowden and Derrick Crook, who have provided much help and inspiration. In particular, I would also like to thank Dona Foster who has been a vital support since the beginning and provided edits and comments throughout my thesis writing, as well as a place to holiday after I submit; Sarah Walker who has been invaluable with statistics help; and Tanya Golubchik who patiently helped me when learning Python.

For the carriage study my thanks go to Kyle Knox who helped with the Ethics application and has submitted every ethical amendment since; Rowena Fung for meticulously carrying out the vast majority of the carriage study lab work; Heather Godwin for recruiting most of the carriage study participants; and Tonya Votintseva for the *spa* typing. For the CC22 CC30 study, thanks to Xavier Didelot for his help with ClonalFrame, amongst advice on many other things; and Danny Wilson, Camilla Ip, Liz Batty and Rosalind Harding for their general support surrounding population genetics and bio-informatics. Also thanks to my team of trusty proof readers, each who have been over at least one chapter at some point: Nancy Miler, Esther Robinson, Hanna Larner, Kate Dingle, Jessica Wray, Sharon Egan, Martin Evans, Louise Glenn, Douglas Russell, Ruth Pidsley, Ellen Miller and Rowena Fung. Finally thanks to my family: Nancy, Bill, Paul and Ellen Miller; who have provided a safe place to turn to every time it got too much and supported me throughout.

Thanks also to my examiners Nick Day and Jodi Lindsay.

Funding Bodies

The work towards this thesis was supported by both the Oxford NIHR Biomedical Research Centre and the UKCRC Modernising Medical Microbiology Consortium, with the latter being funded by the MRC (G0800778) and the Wellcome Trust (087646/Z/08/Z).

Publications and Conference Attendances

Publications

Miller, R; Walker, AS; Knox, K; *et al*. (2010) “Feral” and “wild” type *Staphylococcus aureus* in the United Kingdom. *Epidemiology and Infection*. 2010 May; 138(5) 665-65

Miller, R; Esmail, H; Peto, T; Walker, AS; *et al.* (2008) Is MRSA admission bacteraemia community-acquired? A case control study. *Journal of Infection.* 2008 Mar; 56(3) 163-70

Presentations

Miller, R; Golubchik, T; Wilson, D; *et al.* (2011) Diversity of Two *Staphylococcus aureus* Clonal Complexes (CCs) Over Space and Time. ASM, New Orleans

Miller, R; Golubchik, T; Wilson, D; *et al.* (2010) Diversity of *Staphylococcus aureus* Clonal Complexes (CCs) over space and time. MEEGID, Amsterdam

Posters

Miller, R; Golubchik, T; Wilson, D; *et al.* (2010) Analysis of relationships within a *Staphylococcus aureus* clonal complex (CC) using whole-genome sequencing. ICAAC, Boston

Miller, R; Knox, K; Godwin, H; *et al.* (2010) Carriage of *Staphylococcus aureus* in Oxfordshire, UK. ICAAC, Boston

Miller, R; Walker, S; Knox, K; *et al*. (2008) Methicillin Resistant *Staphylococcus aureus* (MRSA) Acquired Outside Acute Hospital in Oxfordshire. ICAAC, Washington DC

Miller, R; Walker, S; Knox, K; Mant, D; *et al*. (2008) Ciprofloxacin Sensitive (CipS) Methicillin Resistant *Staphylococcus aureus* (MRSA) as a Marker for Community-Acquired (CA-)MRSA. ICAAC, Washington DC

Miller, R; Esmail, H; Peto, *et al*. (2006) Clinical and molecular profile of Methicillin resistant *S. aureus* bacteraemia (MRSA) on admission to a UK hospital. FIS, Cardiff

Attributions

Chapter Three

All data collection, laboratory, statistical and genetic analysis was performed by myself unless stated. For the admission bacteraemia study Dr Hanif Esmail acted as an independent assessor of the details taken from medical records, whilst I collected the information. For the community-cases I sent the faxes and contacted the GPs and Dr Kyle Knox went to GP practices to investigate the records of the eight individuals with no risk factors or only antibiotic risk factors. Elizabeth Howarth confirmed that ethical approval was not required for the community-case and CipS study. Dr John Paul collected epidemiological information for the Brighton CipS cases. I conducted all the statistical analysis under the supervision of Dr Sarah Walker.

Chapter Four

I designed the study with the help of my supervisors and wrote the ethical application under the supervision of Dr Kyle Knox. I attended the ethics board review. Further ethical amendments since the original approval were applied for by Dr Kyle Knox. I wrote the participant information sheets, consent forms and questionnaires. I created the access database and study website. Heather Godwin recruited the majority of patients, however I recruited a small minority to understand this process. I designed the protocol for *S. aureus* isolation, transportation, culture and identification from nasal swabs. I conducted all of the laboratory for work the first three months. Thereafter the majority of the work was performed by the research assistant Rowena Fung. I set up *spa* typing in Oxford including establishing the software. I trained Dr Antonina Votintseva on its use and thereafter she did the majority of *spa* typing including designing the protocols for *spa* typing using the robot and identification of multiple *spa* types from a single nasal swab. I extracted the data from the database and conducted all the statistical analysis myself under Dr Sarah Walker's supervision.

Chapter Five

I designed the study with the help of my supervisors. I extracted the DNA, which was submitted to the sequencing facility. The sequence data was returned as raw fastq files. I performed the initial mapping using Maq and designed the initial quality based filter. Under Dr Daniel Wilson's supervision I developed the goodness of fit filter and wrote the Python scripts to process the data.

Later bioinformatics analysis, including mapping using Stampy and the MMM pipeline filter was performed by the MMM pipeline team including Dr Liz Batty and Dr Camilla Ip. All further analysis was performed using bespoke Python scripts written by myself (except where highlighted: Dr Madeline Cule wrote the pairwise difference and heat map scripts and Dr Liz Batty performed *de novo* assembly and subsequent analysis for SCCmec typing). Dr Xavier Didelot made the changes required to use ClonalFrame on whole-genome sequence data. I performed the statistical genetic analysis (e.g. π_{xy}) and processed the output into appropriate depictions contained in the figures.

Glossary of Abbreviations

agr	Accessory gene regulator
AIC	Akaike Information Criterion
ATP	Adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
bp	Base pair
BP	Blood pressure
BURP	Based upon repeat pattern
BURST	Based upon Related Sequence Types
BWA	Burrows-Wheeler Aligner
CA	Community-acquired
CBA	Columbia Blood Agar
CC	Clonal Complex
CDC	Centres for Disease Control
CI	Confidence interval
CipR	Ciprofloxacin resistant
CipS	Ciprofloxacin sensitive
COPD	Chronic obstructive pulmonary disease
DHAP	Dihydroxyacetone phosphate
d_N	Number of non-synonymous substitutions
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide
d_S	Number of synonymous substitutions
EDTA	Ethylenediaminetetraacetic acid
egc	Enterotoxin gene cluster
ET	Exfoliative toxins
G	Gravity
GP	General Practitioner
HA	Healthcare-acquired
HCl	Hydrochloric acid

HGT	Horizontal gene transfer
IQR	Interquartile range
kb	Kilo base
L	Litre
LCI	Lower confidence interval
LQ	Lower quartile
M	Molar
MAU	Medical assessment unit
mb	Mega base
MCMC	Markov chain Monte Carlo
MFP	Multinomial fractional polynomial
mg	Milligram
$MgCl_2$	Magnesium chloride
MGE	Mobile genetic element
MIC	Minimum inhibitory concentration
min	Minute
ml	Millilitre
MLST	Multi-locus sequence type
mm	Millimetre
mM	Millimolar
MMM	Modernising Medical Microbiology
MREC	Multi-Centre Research Ethics Committee
MRSA	Meticillin-Resistant *Staphylococcus aureus*
MSA	Manitol Salt Agar
MSSA	Meticillin-Sensitive *Staphylococcus aureus*
NaCl	Sodium Chloride
NCBI	National Centre for Biotechnology Information
NEQAS	National External Quality Assessment Service
ng	Nanogram
NHANES	National Health and Nutrition Examination Survey

NHS	National Health Service
ONS	Office of National Statistics
OR	Odds ratio
ORH	Oxford Radcliffe Hospitals
PBP	Penicillin binding protein
PCR	Polymerase chain reaction
PFGE	Pulsed field gel electrophoresis
PVL	Panton Valentine Leukocidin
r/m	Ratio of the number of base changes due to recombination to base changes due to mutation
RAPD	Random Amplification of Polymorphic DNA
RRR	Relative risk ratio
s	Second
SaPI	*Staphylococcus aureus* pathogenicity island
SE	Staphylococcal enterotoxin
SLV	Single locus variant
SNP	Single nucleotide polymorphism
SOP	Standard operating procedure
ST	Sequence Type
STARS	Sequence Typing Analysis and Retrieval System
TBE	Tris Borate Ethylenediaminetetraacetic acid
TE	Tris Ethylenediaminetetraacetic acid
TSST	Toxic shock syndrome toxin
UCI	Upper confidence interval
UKCRC	UK Clinical Research Collaboration
UPGMA	Unweighted Pair Group Method with Arithmetic Mean
UQ	Upper quartile
WTCHG	Wellcome Trust Centre for Human Genetics
ρ/Θ	Ratio of recombination events to mutation events
µg	Microgram

µl	Microlitre
µM	Micromolar

Table of Contents

1 Introduction

1.1 *Staphylococcus aureus*

Staphylococcus aureus is a facultative anaerobic gram-positive coccus, with a characteristic golden colour, often termed “golden staph”. It was discovered by Alexander Ogston in 1880 from pus in a surgical abscess (Ninth Surgical Congress, Berlin, 1880; translated in (Ogston 1984)). In recent years *S. aureus* has become an increasing problem, largely due to its ability to acquire resistance to numerous antibiotics, including meticillin. The resulting meticillin-resistant *S. aureus* (MRSA) has since become infamous in the media where it has been called a “superbug that heralds the end of the golden age of medicine” (Washer and Joffe 2006).

Standard identification of *S. aureus* in the laboratory has classically involved a gram strain, which shows typical gram positive bacteria growing in clusters; and testing for presence of catalase (production of bubbles in hydrogen peroxide); coagulase (coagulation in rabbit plasma); DNAse (zone of clearance on nutrient agar); and Protein A (latex agglutination test) (Health Protection Agency 2007). More recently chromogenic nutrient agar, which detects phosphatase activity present in *S. aureus* strains causing it to develop a distinct colour, has been used in routine practice for *S. aureus* identification. Additionally, salt broth (5% sodium chloride (NaCl)) is selective for *S. aureus* and may be used to enrich samples for *S. aureus,* particularly before growth on chromogenic agar. Molecular techniques, such as the polymerase chain reaction (PCR), can also be used for *S. aureus* identification. PCR that works directly from clinical samples has the advantage of producing results within hours, rather than the days taken for bacterial colonies to grow (Huletsky, Giroux et al. 2004).

S. aureus is a leading cause of hospital morbidity and mortality worldwide, as well as a harmless commensal carried by approximately one third of the global population. The relationship between the bacterial coloniser and human host is multifaceted; and includes host health, socio-demography, and immunity as well as bacterial characteristics such as pathogenicity and antibiotic resistance. Despite much research surrounding the above factors, leading to many advances, including development of new antibiotic therapies and improvements in hygiene, the burden of *S. aureus* disease remains high, and in part unexplained. This thesis will therefore

focus on two areas of *S. aureus* that are poorly understood, in order to further explain the continued threat *S. aureus* poses to human health. Firstly, the reservoir of *S. aureus* in the community and its relationship to *S. aureus* in hospital is studied (Chapters Three, Four and Five); and secondly, the whole-genomic sequence diversity of two *S. aureus* lineages is investigated for evidence explaining their epidemic properties (Chapter Six).

1.2 Pathology

S. aureus is largely a harmless commensal, carried asymptomatically. The primary site for *S. aureus* carriage is the anterior nares; however *S. aureus* can also be isolated from other anatomical sites including the throat, axilla, groin, perineum, abdomen and other areas of skin (Williams 1963; Sanford, Widmer et al. 1994; Sivaraman, Venkataraman et al. 2009). Infection with *S. aureus* is often initiated by a skin breach, allowing access to tissue or the blood. Infections range from superficial skin infections to more severe bacteraemia. The most superficial skin infections are localised, such as impetigo, or folliculitis. Deeper skin infections caused by *S. aureus* include carbuncles and furuncles, which may lead to more serious infections (Lowy 1998; Mandell 2005) including infection of the bone or bone marrow in osteomyelitis, or of the joint in arthritis (Strohl 2001).

More serious infection occurs when *S. aureus* enters the blood which may seed infection in other parts of the body, leading to conditions such as pneumonia, meningitis, endocarditis, and septicaemia (Strohl 2001; Mandell 2005). A worldwide study of over 15000 *S. aureus* bloodstream isolates found *S. aureus* to be the most common cause of bloodstream infection, skin and soft tissue infection and pneumonia in almost all geographic areas (Diekema, Pfaller et al. 2001). *S. aureus* can also cause toxin mediated disease, of which toxic shock syndrome is the best known, as well as staphylococcal scalded skin syndrome and food poisoning (Strohl 2001).

1.2.1 Mortality

Mortality rates from *S. aureus* vary, as demonstrated in a meta-analysis of *S. aureus* bacteraemia where mortality rates in 31 studies varied from 0% to 83% (Cosgrove, Sakoulas et al. 2003). In the John Radcliffe Hospital, Oxford, where research for this thesis was carried out, *S. aureus* bacteraemia had a mortality rate of 29% in 1997-2003 (Wyllie, Crook et al. 2006). Various studies have shown that MRSA bacteraemia has higher mortality than Meticillin-Sensitive *S. aureus*

(MSSA) bacteraemia, with median mortality rates from each of two studies of 34% and 21% for MRSA and 25% and 13% for MSSA (Cosgrove, Sakoulas et al. 2003; Grundmann, Aanensen et al. 2010), however not all the literature agrees with these findings. This may be because features associated with increased *S. aureus* mortality such as greater age and other co-morbidities (Lowy 1998) are also risk factors for MRSA.

Recently, presentation and outcome of *S. aureus* disease has changed due to increasing levels of community-acquired (CA)-MRSA (section 1.4.3). These CA-MRSA infections often arise without obvious skin breach, and the majority (75-80%) present as skin and soft tissue infection, compared to only 37% of nosocomial MRSA (Naimi, LeDell et al. 2003; Davis, Perri et al. 2007).

1.3 Virulence determinants

S. aureus disease presentation is complex, controlled in many cases by a combination of several different genes as well as environmental or host factors. It is therefore difficult to find associations between single genes and *S. aureus* disease presentation. *S. aureus* virulence may be associated with strain or lineage if particular virulence genes or groups of genes are found within one lineage but not another (Diep, Carleton et al. 2006). For example, clonal complex (CC)45 (as defined by Multi-locus Sequence Typing (MLST) (section 1.7.2)) is significantly more likely to be involved in carriage than disease (Melles, Gorkink et al. 2004). However, not all studies have found associations between *S. aureus* presentation and lineage as defined by MLST (Day, Moore et al. 2001; Day, Moore et al. 2002; Feil, Cooper et al. 2003).

1.3.1 Putative virulence factors

Despite the difficulty in finding strong associations between individual genes and disease presentation, many putative *S. aureus* virulence factors have been identified. Virulence factors can be categorised into two groups: toxins and adhesins. Toxins include: enterotoxins released into the intestine causing food poisoning; exfoliative toxins, which lead to exfoliation of the dermis; and haemolysins which cause lysis of red blood cells. Adhesins are required for *S. aureus* to adhere to cells in the human body, and are therefore required for colonising a new host. Examples of commonly referred to enterotoxins and adhesins are outlined in table 1.1.

Table 1.1: Putative *S. aureus* virulence factors

Virulence Determinant	Details	Reference
Toxins:		
Panton-Valentine Leukocidin (PVL)	Pore forming toxin acting on white blood cells encoded by two genes: *lukF-PV* and *lukS-PV* Associated with CA-MRSA Associated with *S. aureus* causing abscesses and arthritis	(Melles, Gorkink et al. 2004; Ferry, Perpoint et al. 2005; Tristan, Ferry et al. 2007)
Enterotoxin gene cluster (*egc*)	Located on the genomic island vSAß Contains staphylococcal enterotoxins (SE)G, SEI, SEM, SEN, and SEO Various studies have found *egc* to be related to carriage rather than disease	(van Belkum, Melles et al. 2006; Nashev, Toshkova et al. 2007)
Exfoliative toxins	Exfoliative toxins (ET)A, ETB and ETD cause exfoliation of the epidermis ETB causes Staphylococcal scalded skin syndrome, often seen in children with peeling of the skin	(Ferry, Perpoint et al. 2005).
Haemolysins	Include α-toxin, β-toxin, and γ-toxin encoded by *hla*, *hlb* and *hlg* respectively	(Projan 1997)
Toxic shock syndrome toxin (TSST-1)	Superantigenic toxin encoded by the *tst* gene *S. aureus* positive for *tst* is highly virulent and causes a variety of illnesses, ranging from toxic shock syndrome to various pus related infections	(Tristan, Ferry et al. 2007)
Adhesins:		
Protein A (*spa*)	Major component of *S. aureus* cell wall Binds to human antibodies and impedes phagocytosis	(Gomez, O'Seaghdha et al. 2006)
Fibronectin-Binding Proteins (*FnbpA* and *FnbpB*)	Adhesin for fibronectin Promotes binding to mucosal cells and tissue matrices	(Peacock, Moore et al. 2002) (Strohl 2001)
Fibrinogen-Binding Proteins (*ClfA* and *ClfB*)	Adhesion for fibrinogen Similar to *FnbpA* and B, however, *clfA* mutants have a greater impact on attachment	(Projan 1997)
Collagen Adhesin (*cna*)	Produces collagen binding protein	(Projan 1997)
Sdr family (*sdrC*, *sdrD*, *sdrE*)	Putative adhesins with unknown function	(Peacock, Moore et al. 2002)
ica locus	Produces biofilm	(Projan 1997)

Studies investigating the association of individual virulence genes with *S. aureus* disease presentation have had limited success. For example, investigation of 33 putative virulence genes in 334 *S. aureus* isolates from healthy blood donors and individuals with invasive disease found seven genes associated with virulence: *fnbA*, *cna*, *sdrE*, *sej*, *eta*, *hlg*, and *ica* (Peacock, Moore et al. 2002). However, other genes with well documented disease association, such as *tst* and *etb* showed no association with virulence.

One of the most commonly investigated *S. aureus* virulence determinants, due to its association with community-acquired disease, is the PVL toxin (Millar, Loughrey et al. 2007) (Table 1.1). Incidence of PVL in MRSA is relatively low, comprising 1.6% of all UK MRSA isolates in 2005 (Holmes, Ganner et al. 2005). However, the prevalence is significantly higher in *S. aureus* causing abscesses and arthritis, where PVL has been found in 39% of strains, compared to only 2% of strains isolated from blood culture (Melles, Gorkink et al. 2004). Incidence is also higher in CA-MRSA, and some definitions actually require presence of PVL for an isolate to be defined as community-acquired (Holmes, Ganner et al. 2005).

The role of PVL in *S. aureus* disease is still uncertain. Studies looking at outcome of patients with *S. aureus* with and without PVL have showed that PVL is strongly associated with skin and soft tissue infections such as abscesses, skin lesions, boils (furuncles) (Holmes, Ganner et al. 2005) and necrotizing pneumonia (Vandenesch, Naimi et al. 2003). However, other studies have concluded that PVL only causes virulence in conjunction with other virulence factors, such as those outlined in table 1.1 (Voyich, Otto et al. 2006; Diep, Palazzolo-Ballance et al. 2008).

1.3.2 *S. aureus* Pathogenicity Islands (SaPIs)

Many *S. aureus* virulence determinants are located on SaPIs: mobile genetic elements between 14 and 17 kilo bases (kb) in size. SaPIs are able to move between different *S. aureus* strains and between species. By 2010 16 SaPIs had been sequenced. These encoded many putative virulence factors, including staphylococcal enterotoxins and TSST-1 (Lindsay and Holden 2006; Malachowa and DeLeo 2010).

1.4 Epidemiology

1.4.1 Incidence and risk factors for *S. aureus* carriage

Prevalence of *S. aureus* carriage in the healthy community is around one third (Kluytmans, van Belkum et al. 1997; Van den Bergh, Yzerman et al. 1999). Two reviews, looking at 18 and 14 studies of 13873 and 7405 healthy individuals, found rates of carriage to vary between 14% and 64% with a mean carriage rates of 37% and 30%, respectively (Kluytmans, van Belkum et al. 1997; Van den Bergh, Yzerman et al. 1999). Reported rates of *S. aureus* carriage have remained relatively stable over time: a review from 1963 reported similar carriage rates from 16 studies of between 30% and 50% (Williams 1963). Studies in the UK have estimated *S. aureus* carriage

rates of 30% in 99 mothers of new born infants (Peacock, Justice et al. 2003), 23% in 280 members of the general population in Birmingham (Abudu, Blair et al. 2001) and 26% in 46 British attendees to the 2003 ECCMID conference (Nulens, Gould et al. 2005).

Rates of MRSA carriage vary more between sub-populations than *S. aureus* carriage. MRSA rates from individuals in hospital or with healthcare risk factors are higher than those in the community (Salgado, Farr et al. 2003; Davis, Stewart et al. 2004). A meta-analysis of 57 studies looking at MRSA in the community established a pooled estimate of MRSA colonisation of 1.3%. Samples taken in hospital were 2.35 times more likely to carry MRSA than those taken in the community, and when looking only at studies of individuals with no prior healthcare contact they found an MRSA colonisation rate of 0.2% (Salgado, Farr et al. 2003). In the USA the National Health and Nutrition Examination Survey (NHANES) of 9622 people in 2001-2 found an MRSA colonisation rate of 0.8% (Kuehnert, Kruszon-Moran et al. 2006).

To date, previous studies investigating risk factors for *S. aureus* carriage have collected only limited data on covariates, with two possible effects: (i) missing further risk factors; and (ii) failing to identify confounders. For example, two large studies using data from NHANES in the USA investigated carriage in 9622 individuals, of which 2964 (32%) had *S. aureus.* However, the sample was not specifically designed to investigate *S. aureus* so adequate information on potential *S. aureus* risk factors was not gathered, for example household contacts and employment were not investigated, and information on healthcare was limited. Despite this, the studies concluded that age, male gender, ethnicity, lower education and asthma were associated with *S. aureus* carriage (Graham, Lin et al. 2006; Kuehnert, Kruszon-Moran et al. 2006) (Table 1.2). Other similar but smaller studies have also concluded that male gender, younger age and ethnicity are associated with *S. aureus* carriage. Findings from NHANES and other studies investigating *S. aureus* nasal carriage are outlined in table 1.2.

Table 1.2: Risk factors for *S. aureus* carriage from current literature

Study size	Participants	Number swabs	Significant Risk factors identified	Reference
9622	NHANES, USA	1	Age<65, male, less than high school education, asthma, race other than black persons	(Graham, Lin et al. 2006)
9622	NHANES, USA	1	Age 6-11, male, non-hispanic white ethnicity	(Kuehnert, Kruszon-Moran et al. 2006)
823	Community members from Manhattan, USA	1	Male gender	(Miller, Cook et al. 2009)
500	School and university students and employees, Lebanon	1	Younger age, male, contact with healthcare workers, using needle injections, asthma. Protective effect from: twice daily washing nose with water, use of nasal spray, acne	(Halablab, Hijazi et al. 2010)
699	General Practice (GP) patients with non-infectious conditions at and volunteers from electoral role, Australia	1	Caucasian ethnicity, male, age<59	(Munckhof, Nimmo et al. 2009)
450	University students from north Carolina, USA	3	Younger age, male, not having antibiotics in past month	(Bischoff, Wallis et al. 2004)
4030	Pre-surgery adults, Iowa, USA	1	Obesity, male, cerebrovascular accident. Protective effect from: previous antimicrobials, smoking, older age	(Herwaldt, Cullen et al. 2004)
1044	Volunteers from Chinese military camps, China	1	Age<22, university or higher education, <2 years of service, non-smoking, inactive participation in social events	(Qu, Cui et al. 2010)

1.4.2 Incidence and risk factors for *S. aureus* disease

S. aureus nasal carriage is a well documented risk factor for subsequent *S. aureus* disease (Corbella, Dominguez et al. 1997; von Eiff, Becker et al. 2001; Wertheim, Vos et al. 2004; Safdar and Bradley 2008) where disease is usually caused by the same strain as found in the nose (Williams 1963; von Eiff, Becker et al. 2001; Wertheim, Vos et al. 2004). Studies investigating the proportion of *S. aureus* carriers who progress to disease have largely been carried out on high risk populations, and may therefore over-estimate disease incidence. There are, however indications that between 0.6% and 6.8% of individuals with *S. aureus* progress to disease in one year or less (Corbella, Dominguez et al. 1997; Davis, Stewart et al. 2004; Wertheim, van Leeuwen et al. 2005; Oztoprak, Cevik et al. 2006).

Other risk factors for *S. aureus* bacteraemia include older age, ethnicity, gender, renal and geriatric unit admission, previous hospital exposure, cellulitis, presence of a catheter, skin ulcers and diabetes (Cordova, Heath et al. 2004; Tacconelli, Venkataraman et al. 2004; Das, O'Connell et al. 2007; Wyllie, Walker et al. 2007). Interestingly, findings surrounding the effect of gender on *S. aureus* carriage are not consistent, since previous studies have found both male (Tacconelli, Venkataraman et al. 2004) and female gender (Cordova, Heath et al. 2004) to increase likelihood of MRSA bacteraemia. Risk factors for less serious disease, such as skin and soft tissue infection are similar, and include age, ethnicity and healthcare exposure (Miller, Cook et al. 2009).

Risk factors for carriage of meticillin-resistant versus meticillin-sensitive *S. aureus* are unknown as all studies investigating risk factors for meticillin resistance focus on strains causing disease. *S. aureus* causing disease is more likely to be resistant to meticillin than *S. aureus* that is carried, with levels of meticillin resistance calculated from samples of disease causing strains ranging from 25% to 46%. Risk factors for infection with MRSA versus MSSA are older age, previous MRSA, previous hospitalisation, surgery, employment as healthcare worker, pressure sores, presence of a catheter, female sex, antimicrobial use, nursing home residence, tube feeding, and not using injection drugs (Goetz, Posey et al. 1999; Roghmann 2000; Graffunder and Venezia 2002; Lodise, McKinnon et al. 2003; Manzur, Vidal et al. 2007). Although these risk factors are almost entirely from studies of *S. aureus* disease, a study including 146 healthy carriers as well as individuals with disease concluded that risk factors did not differ between the two (Goetz, Posey et al. 1999).

1.4.3 Nosocomial versus community-acquired *S. aureus*

S. aureus in the UK community is primarily meticillin-sensitive, with less than 1% of MRSA isolates in the UK described as community-acquired (Elston and Barlow 2009). In contrast, *S. aureus* acquired in hospitals is much more likely to be meticillin-resistant. In 2006 38% of *S. aureus* bacteraemia reported from UK hospitals was MRSA (Health Protection Agency 2007). An increased incidence of CA-MRSA is also detected in individuals with non-hospital healthcare risk factors, for example district nurse patients have a higher MRSA colonisation rate of 21% (Thomas, Karas et al. 2007) and constitute a significant proportion (7%) of individuals with community-onset MRSA (Bygott, Enoch et al. 2008). Similarly, residents of nursing homes are

known to have higher rates of MRSA carriage, with colonisation rates of 17%, 9% and 10% reported (Fraise, Mitchell et al. 1997; O'Sullivan and Keane 2000).

Levels of MRSA in the UK general community remain unknown. There have been isolated reports of MRSA spreading in the UK non-healthcare community (Holmes, Ganner et al. 2005; David, Kearns et al. 2006; Adedeji, Weller et al. 2007; Otter and French 2008) suggesting that the UK may be moving toward the trend seen in the USA where CA-MRSA has become increasingly common since the 1990s (Groom, Wolsey et al. 2001; Hussain, Boyle-Vavra et al. 2001; Chambers and Deleo 2009).

The epidemiology of *S. aureus* differs between community and healthcare acquired strains and current literature also distinguishes between the two using multiple factors including strain genetics, epidemiology, resistance to antibiotics and disease presentation (Table 1.3).

Table 1.3: Factors distinguishing between healthcare-acquired (HA) and CA- MRSA

Feature	HA-MRSA	CA-MRSA	Reference
Strain characteristics:			
Genotypes	Homogeneous	Heterogeneous	(Grundmann, Aires-de-Sousa et al. 2006)
Staphylococcal cassette chromosome (SCC)*mec* types	Mainly larger types I, II and III	Mainly smaller types IV and V	(Grundmann, Aires-de-Sousa et al. 2006)
agr (accessory gene regulator) type	Predominantly II	I and III	(Millar, Loughrey et al. 2007)
Presence of PVL	Low <5%	High >5%	(Millar, Loughrey et al. 2007)
Resistance pattern	Resistant to multiple antibiotics	Susceptible to most antibiotics	(Grundmann, Aires-de-Sousa et al. 2006)
Epidemiology:			
Isolated from	Individuals with healthcare contact	Closed populations such as: indigenous populations, homeless, men who have sex with men, prisoners, military, children in day care centres, athletes	(Grundmann, Aires-de-Sousa et al. 2006)
Age	Older	Younger	(Diederen and Kluytmans 2006)
Ethnicity	White	Non-white	(Diederen and Kluytmans 2006)
Presentation:			
Disease presentation	Respiratory tract, urinary tract and bloodstream infections	Skin and soft tissue infections	(Diederen and Kluytmans 2006)

1.4.4 Definitions of CA-MRSA

CA-MRSA is distinct from nosocomial or HA-MRSA (Table 1.3) and is often described separately in research and literature. However definitions used to distinguish between CA- and HA-MRSA vary considerably, as highlighted by a meta-analysis of 57 studies investigating CA-MRSA, which found eight different definitions were used (Salgado, Farr et al. 2003). This variety makes it difficult to compare and make inferences from the literature.

It is important to accurately distinguish between CA and HA-MRSA to ensure that instances of CA-MRSA can be properly recognised and treated appropriately. If left untreated CA-MRSA may spread in the community and could become a public health problem. Additionally, proper identification of CA-MRSA means that hospitals can change their policies according to levels of

CA-MRSA or specific populations at higher risk, for example by testing more people on admission to ensure they do not transmit MRSA into the hospital. In addition to this, some evidence suggests that CA-MRSA may be more virulent and more easily transmitted than HA-MRSA (Deleo, Otto et al. 2010), therefore if CA-MRSA were to acquire antibiotic resistance determinants it may be able to outcompete or even replace current epidemic HA-MRSA strains (Millar, Loughrey et al. 2007) leading to new epidemics.

Definitions of CA-MRSA are largely based around three factors: (i) time of isolation; (ii) microbiology; and (iii) epidemiology. Time based definitions are the simplest definitions of CA-MRSA and are often used when the origin of an isolate is uncertain. For example MRSA may be classified as community-acquired if an individual has been in hospital for less than 48 hours before MRSA isolation (Salgado, Farr et al. 2003). However, time based definitions alone are not sufficient to classify MRSA as community-acquired, since they are likely to overestimate MRSA in the community by incorrectly classifying patients with unknown healthcare risk factors as community-acquired. When only time since previous hospitalisation is known, it is best to use the terms community-associated (Millar, Loughrey et al. 2007), or community-onset MRSA which describe the location of the individual when MRSA was identified (Diederen and Kluytmans 2006).

Genetic characteristics used to define CA-MRSA include looking for strains not commonly seen in hospitals, identification of PVL or identification of SCC*mec* type IV, which, although primarily associated with CA-MRSA, is not a unique feature (Charlebois, Perdreau-Remington et al. 2004), since it is also found in the UK epidemic healthcare associated strain EMRSA-15 (Conceicao, Aires-de-Sousa et al. 2007). However, use of genetic characteristics alone does not prove whether strains are actually acquired in the community and as strains genetically similar to those that are community-acquired begin to be seen in hospitals and vice versa use of genetic characteristics to define CA-MRSA is becoming increasingly invalid. For example, strains genetically identical to healthcare-acquired strains may spread in the community in individuals with other healthcare related activities such as help at home, these strains have been termed feral as they represent community isolates of nosocomial origin (Bhattacharya, Carleton et al. 2007).

Empirical definitions of CA-MRSA consider epidemiology, including use of information from case notes and medical history. An example of an epidemiological based definition of CA-MRSA is the Centers for Disease Control (CDC) definition, which considers time of isolation as well as medical history and risk factors (Table 1.4).

Table 1.4: CDC definition of CA-MRSA (Grundmann, Aires-de-Sousa et al. 2006)

MRSA diagnosed in out-patient or within 48 hours of admission to hospital
No medical history of MRSA infection or colonisation
No history in the past year of: • Hospitalisation • Admission to a nursing home, skilled nursing facility, or hospice • Dialysis • Surgery
No permanent indwelling catheters or medical devices that pass through the skin into the body

When classifying a strain as community-acquired it is important to consider all aspects, including epidemiology, previous healthcare, presentation and *S. aureus* genetics, before concluding that it is a true community-acquired or "wild" strain. Millar *et al* have proposed a new rigorous definition of CA-MRSA that takes into account clinical, microbiological and epidemiological characteristics. They suggest that for a strain to be classed as community-acquried it must: (i) be confirmed as MRSA using standard laboratory procedures; (ii) fit the CDC definition for CA-MRSA; (iii) be susceptible to more antibiotics than other healthcare strains; (iv) usually present as a skin and soft tissue infection; and (v) have SCC*mec* type IV or V (Millar, Loughrey et al. 2007). Alternatively one could argue that the classification of MRSA into healthcare and community associated is becoming invalid and is not something we should endeavour to distinguish between at all.

1.5 Carriage Dynamics

1.5.1 Estimates of carriage duration

The duration of *S. aureus* carriage in healthy individuals is unknown. Reported durations of *S. aureus* carriage vary widely: from 70 days in patients who acquired MRSA on admission to hospital (MacKinnon and Allen 2000); to 15 months for MRSA positive patients discharged from hospital (Vriens, Blok et al. 2005); and 40 months for known carriers readmitted to hospital (Sanford, Widmer et al. 1994).

Much of the uncertainty in estimates of carriage duration is due to the large amount of variation in the format of previous studies. Many longitudinal *S. aureus* carriage studies have been carried out at Erasmus University in the Netherlands. Such studies include cohorts of 31 volunteers with 10 nasal swabs taken over a period of 19 months (Van Belkum, Riewarts Eriksen et al. 1997); 4797 elderly participants each with two nasal swabs (van den Akker, Nouwen et al. 2006); 51 volunteers with 12 weekly nasal swabs (Nouwen, Ott et al. 2004), of whom 17 were re-swabbed after eight years (Van den Bergh, Yzerman et al. 1999); and 51 volunteers followed for a median of six months, then treated with mupirocin and re-swabbed (van Belkum, Verkaik et al. 2009). Each of these studies, as well as those from other groups, uses different sample sizes, follow-up lengths and frequencies of swabbing; and either follows participants for short time periods or with infrequent testing, making it hard to get an accurate and comparable picture of longitudinal *S. aureus* carriage. Additionally, many studies investigating *S. aureus* carriage follow patients previously known to be MRSA positive so the findings may not be extrapolated to MSSA carriage.

Many studies classify individuals as intermittent or persistent carriers, as an alternative to investigating carriage duration. It is thought that a subset of individuals, about 15-35%, carry *S. aureus* persistently (Van Belkum, Riewarts Eriksen et al. 1997; Van den Bergh, Yzerman et al. 1999; van den Akker, Nouwen et al. 2006). However, definitions of carrier states are highly varied. Two reviews of 14 studies in total showed that definitions of persistent carriage vary from requiring 80-100% of samples to be positive, and give estimates of 9-37% of individuals being *S. aureus* persistent carriers (Williams 1963; Van den Bergh, Yzerman et al. 1999). The longest follow-up period in any study reviewed was 19 months so if persistent carriers carry for longer than this they may not have been properly identified, even if definition required 100% of swabs to be positive. Additionally, Van den Bergh and colleagues reported a study where 17 individuals who had been persistently positive on weekly nasal swabs for 12 weeks, were re-swabbed eight years afterwards. Here 12 participants (71%), still carried *S. aureus*, of whom 25% carried the same strain as defined by PFGE, suggesting that true persistent carries may continue to carry *S. aureus* for highly prolonged periods (Van den Bergh, Yzerman et al. 1999).

It has been suggested that, if properly identified with sufficient follow-up duration and frequency of swabbing, persistent carriers are a different group from intermittent and non-carriers (van Belkum,

Verkaik et al. 2009). Persistent carriers may carry *S. aureus* for many years (Van den Bergh, Yzerman et al. 1999) and have been demonstrated to behave differently from intermittent and non-carriers in controlled experiments. In an experiment performed at Erasmus University where *S. aureus* carriage was cleared with mupirocin, following which individuals were re-infected with various *S. aureus* strains, persistent carriers carried for longer, with more colonies, and were significantly more likely to re-acquire the strain they carried before than intermittent and non-carriers who behaved similarly to each other (van Belkum, Verkaik et al. 2009). This led to the conclusion that there are two carriage types: persistent and other, as non-carriers may just be a sampling artefact and all individuals who do not carry persistently may carry intermittently in the right environment.

Studies correlating host genetic factors with longitudinal carriage further support that persistent carriers are a distinct group. Two studies from Erasmus University using cohort of approximately 3000 individuals, each with two nasal swabs taken at least seven years apart, compared persistent carriers: individuals positive for *S. aureus* on both nasal swabs, with intermittent carriers who were *S. aureus* positive on only one swab. From these studies associations were found between polymorphisms in glucocorticoid receptor genes and interleukin-4 with persistent carriage; as well as polymorphisms in C-reactive protein with non-carriage (van den Akker, Nouwen et al. 2006; Emonts, Uitterlinden et al. 2008). Interestingly, significant host genetic polymorphisms were not found between intermittent and persistent carriers, only persistent carriers versus others (Emonts, Uitterlinden et al. 2008). However, the simple classification of persistent and intermittent carriage from two nasal swabs, which were not typed to determine whether the strain has changed, may miss associations that would have arisen if persistent carriage was defined more rigorously.

1.5.2 Epidemiological factors affecting carriage duration

Few studies investigating *S. aureus* carriage have considered both demographic risk factors and carriage duration in order to investigate factors affecting carriage length; and those that have only investigated MRSA, and often had short durations, making it difficult to identify true persistent carriers. Four studies investigating risk factors for *S. aureus* carriage duration followed participants for between 30 days and 4.4 years, each correlating participant risk factors with slightly different outcome variables (Table 1.5). Of these four studies only one typed strains to

investigate whether carriage changed over time and even then did not correlate change in PFGE type with risk factors collected (Scanvic, Denic et al. 2001).

Table 1.5: Characteristics of studies investigating risk factors for increased *S. aureus* carriage duration

Study Duration	Participants	Outcome variable	Risk factors for increased likelihood of outcome variable	Reference
10 months	Previous MRSA carriers readmitted to hospital	Carriage remaining after three months	Skin breaks	(Scanvic, Denic et al. 2001)
One year	People living in care facilities	MRSA clearance within a year	Self sufficiency in daily activities	(Lucet, Paoletti et al. 2009)
30 days	Patients in hospital with known MRSA carriage	MRSA remaining for 30 days	Female, recent fluoroquinolones, more colonised bodily sites	(Harbarth, Liassine et al. 2000)
Mean 16.2 months	Patients with first time MRSA detected in hospital	MRSA clearance	Antibiotics, indwelling devices, skin lesions, immunosuppressive therapy, haemodialysis	(Marschall and Muhlemann 2006)

1.5.3 Carriage of more than one *S. aureus* strain

S. aureus strains carried in the nose may change over time, however few studies investigating carriage duration consider this. Carriage duration is often measured at the level of the individual, where they are classified by presence or absence of *S. aureus* in the nose, however, consideration at the level of the bacteria, may reveal multiple *S. aureus* strains changing over time. A study following carriage over eight years found 12/17 individuals still carried *S. aureus*, but only three had the same type as defined by Random Amplification of Polymorphic Deoxyribonucleic acid (RAPD) and PFGE (Van den Bergh, Yzerman et al. 1999). Therefore, consideration of type may significantly reduce estimates of carriage duration.

Investigation of carriage duration at the strain level may also find multiple strains in the nose simultaneously. A mathematical model developed after finding 14/149 individuals carrying *S. aureus* harboured more than one strain at a time concluded that 6.6% of *S. aureus* colonised individuals carry more than one strain (Cespedes, Said-Salim et al. 2005). Thus if carriage duration is measured at the strain level, carriage of multiple strains may be underestimated if only one colony per sample is investigated.

1.5.4 Bacterial genetics associated with carriage duration

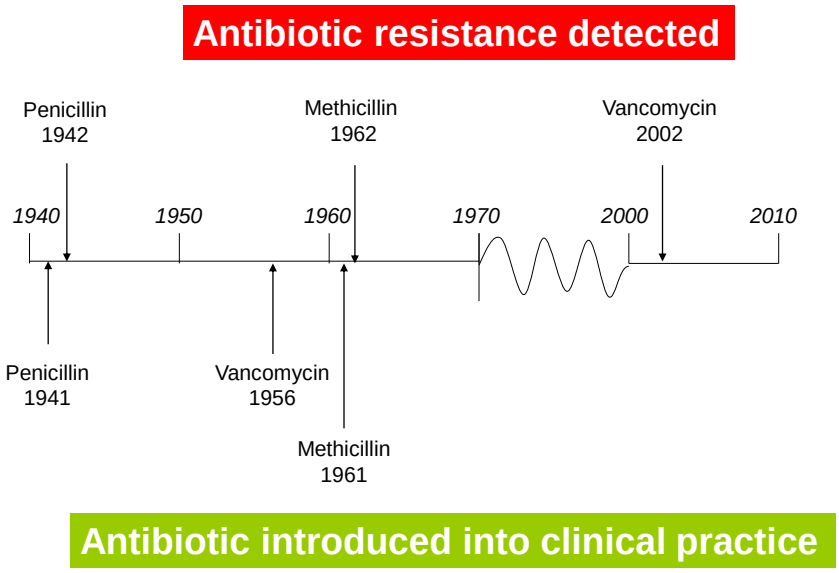
No studies have found correlations between strain of *S. aureus* carried and carriage duration or type, using various techniques including MLST, PFGE (Peacock, Justice et al. 2003), phage typing and RAPD (Van Belkum, Riewarts Eriksen et al. 1997). It has been postulated that protein A (which includes the polymorphic X region sequenced in *spa* typing) may be involved in adherence of *S. aureus* to host cells (Haagen, Heezius et al. 1990) and therefore related to carriage duration. As yet, this has not been demonstrated (Van Belkum, Riewarts Eriksen et al. 1997; Van den Bergh, Yzerman et al. 1999), however studies investigating this had small sample sizes and investigated the length rather than the actual sequence of the protein A gene so it is possible that a relationship between *spa* type and carriage duration is yet to be demonstrated.

1.6 Antibiotic resistance

S. aureus is notable for its ability to rapidly require resistance to antibiotics. Nosocomial *S. aureus* is subject to a large selective pressure from antibiotic use and the ability of *S. aureus* to acquire genetic cassettes from other similar species provides an easy way in which strains of *S. aureus* can rapidly develop new forms of antibiotic resistance. The most notable antibiotics *S. aureus* is resistant to are penicillin, meticillin and vancomycin (Figure 1.1), however, this is by no means an exclusive list. *S. aureus* in the UK is most often resistant to penicillin; ciprofloxacin, 45% (samples resistant in the UK); flucloxacillin, 44%; erythromycin 40%; trimethoprim 13%; and fusidic acid 11% (BSAC Bacteraemia Resistance Surveillance Programme. 2006 data, www.bsacsurv.org, accessed 09.05.2008, (Health Protection Agency 2007). Resistance to penicillin is so common that it is not used to treat patients with *S. aureus* empirically and penicillin resistance levels are not measured, however it has been documented that more than 90% of *S. aureus* today is resistant to penicillin (Schito 2006).

Figure 1.1: Timeline of *S. aureus* antibiotic resistance

Events below the timeline indicate when the antibiotic was first introduced into clinical practice and events above the timeline when resistance to each antibiotic was first detected.



1.6.1 Penicillin resistance

Penicillin was introduced in 1941 (Chambers 2001) and penicillin resistant *S. aureus* first identified in 1942 (Lowy 2003) (Figure 1.1) with the first paper documenting this in 1944 (Kirby 1944). By 1948 50% of *S. aureus* strains found in hospital were resistant to penicillin, with levels of resistance rising further to 80% in 1957 (Schito 2006). Resistance to penicillin in *S. aureus* is encoded by *blaZ*, a gene encoding β-lactamase which degrades penicillin before it reaches its target (Oliveira, Tomasz et al. 2002). Four types of *blaZ* have been identified, three of which are usually located on plasmids, and are therefore transferable between species of bacteria (Olsen, Christensen et al. 2006). The plasmid also contains genes encoding resistance to other antibiotics, such as gentamicin and erythromycin.

1.6.2 Meticillin resistance

Meticillin was recognised in 1961 and resistance emerged in less than one year (Chambers 2001). Since then prevalence of *S. aureus* resistant to meticillin has risen worldwide from 2% in 1975 to 45% in 1999 (Diekema, Pfaller et al. 2001). Resistance to meticillin is encoded by the *mecA* gene situated in the SCC*mec* cassette, a horizontally acquired element, originally

transferred to *S. aureus* from another bacteria, possibly *S. sciuri* (Couto, Wu et al. 2003). SCC*mec* ranges in size from 21–67 kb (Okuma, Iwakawa et al. 2002; Hanssen and Ericson Sollid 2006) and integrates into *S. aureus* at a specific site, *att*B, found at the end of the open reading frame, *orf*Xm (Noto, Kreiswirth et al. 2007). SCC*mec* has currently been divided into eight types (Zhang, McClure et al. 2009), with new types and subtypes continuing to be found. The most commonly identified types are I-IV, which are often used for *S. aureus* typing. *mecA* encodes an altered penicillin-binding protein (PBP) called PBP2a which replaces the biosynthetic function of native PBPs whilst being resistant to β-lactam inhibition because of reduced affinity, thus allowing *S. aureus* to survive exposure to all β-lactam antibiotics (Schito 2006). Therefore resistance to meticillin also confers resistance to all other β-lactam antibiotics including cephalosporins (Lowy 2003) such as flucloxacillin, cefalexin and cefuroxime.

1.6.3 Vancomycin resistance

In 1997 the first strains of MRSA with reduced susceptibility to vancomycin were reported (Hiramatsu, Hanaki et al. 1997) and the first fully vancomycin-resistant strain was reported in 2002 (Sievert 2002). Compared to resistance to other antibiotics, vancomycin resistance took a long time to emerge: 41 years from first introduction in 1956 (Chambers 2001) (Figure 1.1). The mechanism for vancomycin resistance differs between strains with reduced vancomycin susceptibility or complete resistance. To become fully vancomycin resistant requires acquisition of *vanA* operons, originating in *Enterococcus faecalis* (Lowy 2003).

1.6.4 Other antibiotic resistance

Resistance to ciprofloxacin and other antibiotics in the quinolone group also emerged soon after their introduction. For unknown reasons quinolone resistance is more common in meticillin-resistant strains (Lowy 2003). Today ciprofloxacin resistance is of increased interest as ciprofloxacin sensitivity in MRSA is used as an indicator of community acquisition (Otter and French 2008). Resistance to quinolones may be caused by many mutations including point mutations in *grlA* and *gyrA,* encoding deoxyribonucleic acid (DNA) topoisomerase and DNA gyrase respectively (Schmitz, Jones et al. 1998), both targeted by quinolones. An additional mechanism causing low-level quinolone resistance involves increased expression of the *norA* multidrug resistance efflux pump (Lowy 2003).

Erythromycin, a macrolide antibiotic, was first introduced in 1952 and, characteristically for *S. aureus*, resistance arose quickly. There are two mechanisms of resistance. The first is mediated by the genes *ermA*, *ermB* and *ermC,* found on the chromosome and on plasmids, which modify ribosomes and prevent macrolides from binding correctly. The second is encoded by the gene *mrsA* which alters an Adenosine triphosphate (ATP) dependent pump, causing it to remove the antibiotic from the cell and maintaining concentrations too low for it to be effective (Schito 2006).

1.7 Molecular typing

Distinguishing between bacterial strains using genetic features (molecular typing) has applications in the short term such as outbreak identification, as well the ability to make inferences over longer timescales about bacterial evolution. Until recently molecular typing was the only method to assess relatedness of strains and infer evolutionary history. Therefore many studies have used molecular typing to investigate *S. aureus*.

1.7.1 PFGE

Until recently PFGE was regarded as the gold standard for *S. aureus* typing (Berglund, Molling et al. 2005). PFGE involves digestion of DNA using *SmaI* restriction endonucleases, producing fragments which are separated by gel electrophoresis in which the voltage of the gel is periodically switched in three directions allowing separation of larger DNA fragments, above 15-20 kb (Trindade, McCulloch et al. 2003). PFGE has high levels of discrimination compared to MLST and is more useful for detecting outbreaks. The application of PFGE is limited by difficulties in reproducibility and inter-laboratory comparison of results based on arbitrary patterns observed in an agarose gel (Cookson, Robinson et al. 2007). However, standardised methods for identifying outbreaks using PFGE, requiring use of a control strain, and carefully described changes in banding patterns of related isolates have been proposed to overcome this problem (Tenover, Arbeit et al. 1995).

1.7.2 MLST

MLST is a widely used technique that characterises isolates by sequencing 400-600 base regions of six to 10 housekeeping genes. It was developed for use on *Neisseria meningitidis* (Maiden, Bygraves et al. 1998) and schemes have since been developed for numerous bacterial species,

including *S. aureus*, *Streptococcus pneumoniae*, *Campylobacter jejuni*, *Haemophilus influenzae*, *Escherichia coli* and *Salmonella enterica* (Maiden 2006). MLST for *S. aureus* was developed by Enright *et al*, and uses seven loci (Enright, Day et al. 2000) described in table 1.6.

Table 1.6: Seven genetic loci used in *S. aureus* MLST

Gene	Function
Carbamate kinase (*arcC*)	Enzyme that makes carbamoyl phosphate, involved in synthesis of pyrimidines
Shikimate dehydrogenase (*aroE*)	Catalyses the reaction of 3-dehydroshikimate to shikimate. Shikimate is a precursor for making amino acids and other metabolites
glycerol kinase (*glp*)	Catalyses the reaction to form glycerol 3-phosphate from ATP and glycerol
guanylate kinase (*gmk*)	Critical enzyme in the biosynthesis of guanosine 5'-triphosphate, a purine nucleotide
phosphate acetyltransferase (*pta*)	Enzyme involved in carbon cycling and energy metabolism. Converts acetyl coA and phosphate into acetate and ATP
triosephosphate isomerase (*tpi*)	Enzyme that allows the utilisation of dihydroxyacetone phosphate (DHAP) in glycolysis, where ATP is produced
acetyl coenzyme A acetyltransferase (*yqiL*)	Enzyme required for lipid metabolism. Controls the flow of carbon as part of a biochemical pathway involved in formation of acids and solvents

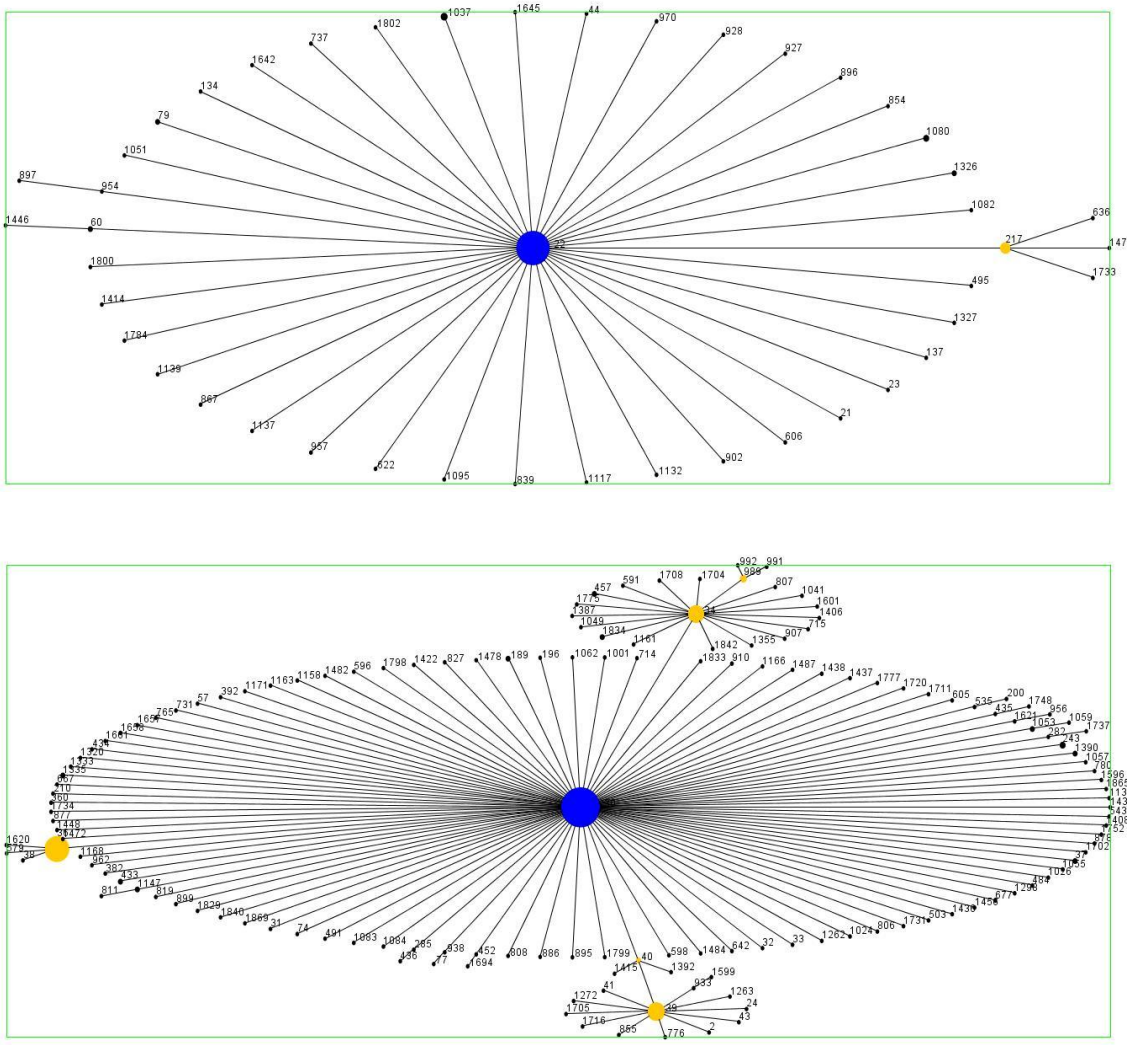
MLST alleles are defined as unique sequences at a particular MLST locus. Sequence types (STs) comprise a unique combination of alleles and thus indirectly define unique sequences at the seven loci. To assign an ST seven allele numbers must first be obtained by entry of DNA sequence for each locus into an internet database (http://saureus.mlst.net). Allele numbers are different for each unique sequence and assigned in the order the alleles are discovered, for example, the first sequence identified for *arcC* in *S. aureus* was assigned *arcC*-1. When all loci from a particular strain have been assigned allele numbers they are combined, using the online database, to produce an allelic profile or ST. For example, one of the most common healthcare-associated *S. aureus* STs is ST22, whose allelic profile is *arcC*-7, *aroE*-6, *glpF*-1, *gmk*-5, *pta*-8, *tpi*-8, *yqiL*-6. Each ST summarises a large amount of data that can be easily compared with strains found worldwide.

Clonal complexes are groups of STs with six out of seven MLST alleles in common with at least one other member of the group, defined by a program called Based upon Related Sequence Types (eBURST) (Feil, Li et al. 2004). The ancestral strain of the CC is allocated by identification of the isolate related to the most other strains in the group by only a single locus variant (Spratt,

Hanage et al. 2004). Figure 1.2 shows eBURST diagrams of two MLST CCs common in UK HA-MRSA: CC22 and CC30.

Figure 1.2: Relationship between STs in CC22 (top) and CC30 (bottom) as defined by eBURST

Blue nodes represent the ancestral strain of the CC and yellow nodes sub-group founders. Numbers are STs.



A limitation of MLST and eBURST analyses is that STs and CCs are not defined directly from the nucleotide sequence so there is no proper distance metric. CCs are calculated using allele numbers and STs rather than nucleotide sequence. This simplifying assumption may be useful in species with high levels of recombination, in which alleles may differ by a large number of nucleotides, due to a single recombination event. However, in clonal species, such as *S. aureus,* which evolve largely through mutation, alleles that differ by more than one nucleotide change are

likely to be more evolutionarily distant than those that differ by only one base change, something that eBURST does not account for (Maiden 2006).

MLST has been used for multiple population studies investigating *S. aureus* (Enright, Robinson et al. 2002; Feil, Cooper et al. 2003; Robinson and Enright 2004). However, it was initially designed for characterising bacteria, so MLST based population studies have been criticised on various grounds. Firstly, seven MLST loci is the minimum required for satisfactory discrimination, but not necessarily the optimum number for population studies, and it might in many cases be better to sequence more loci (Maiden 2006). Additionally, housekeeping genes evolve only very slowly so give information on very distant evolutionary events, but do not discriminate between strains which diverged more recently, such as nosocomial *S. aureus* strains common in UK hospitals. For information on more recent evolutionary events, such an as those within an epidemic, higher resolution techniques are needed. Despite this, phylogenetic trees produced by MLST have highly similar topologies to trees constructed using more genes (Cooper and Feil 2006), suggesting that MLST accurately represents the relationship between isolates. Therefore, until very recently, when the cost and speed of whole-genome sequencing has reduced sufficiently to make population studies using the entire genome a realistic option, MLST was a highly useful method for investigating *S. aureus* evolution.

1.7.3 *spa* typing

spa typing involves sequencing the polymorphic X region of the protein A gene (Harmsen, Claus et al. 2003), which contains between one and 23 variable 21-30 base pair (bp) repeats and is subject to spontaneous mutation. Loss and gain of these repeats results in sufficient variation to distinguish between different *S. aureus* strains. High variability in the number and sequence of repeats allows a typing system where repeats are assigned a numerical code and a *spa* type is assigned from the order and number of specific repeats. The routinely used system for *spa* typing was designed by Harmsen *et al* (Harmsen, Claus et al. 2003) in which sequences are processed using RidomStaphType software (Ridom GmbH, Würzburg, Germany) and the order and number of repeats are used to assign a *spa* type. *spa* types are then uploaded to the internet (http://spaserver2.ridom.de/) where frequencies and details of each are recorded.

spa typing has higher resolution than MLST, for example a study of 99 strains was separated into 22 STs by MLST and 44 *spa* types (Strommenger, Kettlitz et al. 2006) with results correlating well between the two techniques. A study of 59 isolates found a 73% correlation between *spa* types and MLST, where *spa* typing was more diverse (Malachowa, Sabat et al. 2005). That only one gene needs to be sequenced for similar discrimination with higher resolution, gives *spa* typing an advantage over MLST.

spa types can also be assigned to clusters using Based Upon Repeat Pattern (BURP) analysis, which assigns a cost to repeat unit duplication and excision as well as individual point mutations, and then clusters samples related by a cost of four or less (Mellmann, Weniger et al. 2007). BURP clusters correlate well with BURST, with a correlation of 97% demonstrated in a study of 99 *S. aureus* strains (Strommenger, Kettlitz et al. 2006). However, it has been suggested that evolutionary relationships inferred from *spa* typing are not representative of the entire *S. aureus* genome. Similarities in the *spa* gene may not represent a shared evolutionary history of the entire *S. aureus* genome because *spa* types are determined from a relatively small number of repeats, and repeat mutations may occur independently in unrelated strains. In addition, the *spa* gene has been described as a virulence factor (Gomez, O'Seaghdha et al. 2006), suggesting the possibility that convergent evolution driven by host immune selection might obscure evolutionary relationships (Nubel, Roumagnac et al. 2008).

1.7.4 SCC*mec* typing

The SCC*mec* element has moved into *S. aureus* multiple times since the introduction of meticillin, so it does not have the same evolutionary history as the rest of the genome. However, it is useful for supplementing other typing methods. SCC*mec* types are assigned by examining two gene complexes within the element; *mec* and ccr. (Ito, Ma et al. 2004) The *mec* complex contains the *mecA* gene which confers meticillin resistance. The ccr complex is made up of ccr genes; *ccrA*, *ccrB* or *ccrC,* which are site specific recombinases required for movement of the element between bacteria, (Katayama, Ito et al. 2000; Hanssen and Ericson Sollid 2006; Noto, Kreiswirth et al. 2007). Together, *mec* and ccr have been used to divide SCC*mec* into several types, with different structures and functions.

Type I, II and III SCC*mec* are comparatively larger than type IV and V and carry genes for additional antimicrobial resistance. Smaller SCC*mec* elements are more transmissible (Aires de Sousa and de Lencastre 2003) so have become common in CA-MRSA strains that have arisen from meticillin-sensitive isolates in the community acquiring SCC*mec*. Additionally, *in vitro* experiments have demonstrated that carriage of a smaller SCC*mec* element offers a fitness advantage to community strains, which are not subject to the antibiotic pressure of nosocomial isolates (Lee, Ender et al. 2007).

1.7.5 Molecular typing and population biology

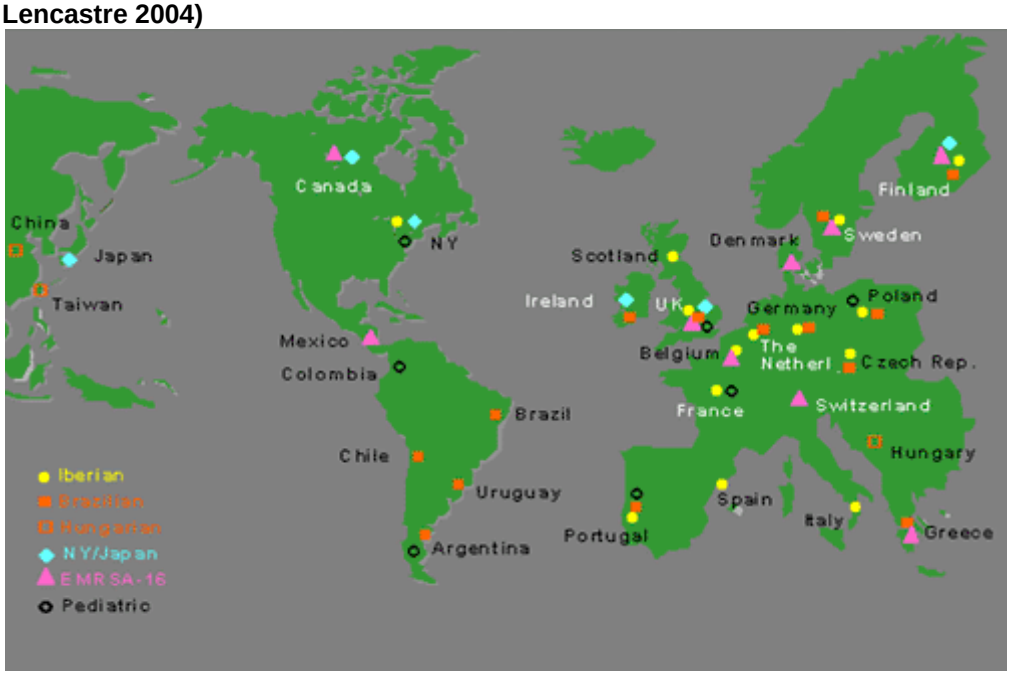
Molecular typing methods have revealed much about *S. aureus* evolutionary history, variability, and virulence. Molecular techniques have been used to compare modern and "archaic" strains to investigate the age of *S. aureus* lineages (Crisostomo, Westh et al. 2001). They have also been used over shorter timescales to identify isolates that are recent descendents or ancestors of common epidemic strains (Hallin, Denis et al. 2007). Additionally, the combination of SCC*mec* typing and other molecular typing schemes has allowed investigation of the evolution of meticillin resistance, including findings that ST250 was the original MRSA clone (Enright, Robinson et al. 2002) and that *S. aureus* may have acquired SCC*mec* more than 20 times (Robinson and Enright 2003; Nubel, Roumagnac et al. 2008).

1.7.6 Molecular typing and strain classification

Understanding the spread and evolution of *S. aureus* requires the existence of secure and transferable typing approaches. MLST, PFGE and SCC*mec* typing have all been used to define nomenclature classifying *S. aureus* strains worldwide. Six pandemic MRSA clones, each distinguishable by a combination of MLST and SCC*mec* typing, have been identified (Figure 1.3) (Aires de Sousa and de Lencastre 2004). They are the Iberian clone: ST247 SCC*mec* I; the Hungarian: ST239 SCC*mec* III; the Brazilian: ST8 SCC*mec* III; the New York/Japan: ST5 SCC*mec* II; the Pediatric: ST5 SCC*mec* IV; and the EMRSA-16 clone: ST36 SCC*mec* II (Enright, Robinson et al. 2002; Oliveira, Tomasz et al. 2002; Conceicao, Aires-de-Sousa et al. 2007; Takano, Higuchi et al. 2007). In the UK the EMRSA-15 clone is also very common, and consists of ST22 with SCC*mec* IV (Conceicao, Aires-de-Sousa et al. 2007). Unfortunately the nomenclature does not always remain consistent between countries, for example EMRSA-16 is the same as USA200 (Holden, Feil et al. 2004). As well as a nomenclature for epidemic MRSA

isolates, common CA-MRSA strains have also received common names. USA300, the cause of most CA-MRSA in the USA, consists of ST8 with SCC*mec* IV (Diep, Gill et al. 2006), and in Europe, a common CA-MRSA strain known as the European clone is defined by ST80 and PVL positivity (Takano, Higuchi et al. 2007).

Figure 1.3: Global distribution of six epidemic MRSA clones (Aires de Sousa and de Lencastre 2004)



1.8 *S. aureus* evolution and diversity

1.8.1 Role of mutation and horizontal gene transfer (HGT) in *S. aureus* evolution

S. aureus is usually described as a clonally evolving organism. This means that evolution occurs more often through single point mutations than recombination. Despite this, HGT is important as a means of generating inter-strain diversity, the majority of which occurs on mobile genetic elements (MGE) including phages, transposons, plasmids and genomic islands. The mechanisms of transfer in bacteria are transformation, transduction, or conjugation, of which, transduction, mediated by bacteriophages, is thought to be the most important in *S. aureus* (Lindsay and Holden 2006). In *S. aureus* natural competence for transformation has not been described and, although conjugative plasmids have been found, they are not thought to contribute much to diversity (Dale and Park 2004; Lindsay and Holden 2006).

Staphylococcal pathogenicity islands (SaPIs), which may carry genes encoding toxins, are transduced by phages and integrate site-specifically into the staphylococcal chromosome.

SCC*mec* is a chromosomal cassette which inserts site-specifically into the genome and carries the meticillin resistance gene *mecA*.

Recombination occurs following HGT if the horizontally transferred DNA is incorporated into the host chromosome. In the case of homologous recombination, where there is homology between the transferred and host DNA, recombination is likely to generate new diversity in a pre-existing allele. Homologous recombination differs from HGT involving elements such as SaPIS or phages, where strains acquire an extra segment of DNA and can occur in stable regions of the genome such as housekeeping genes. Investigation of HGT in *S. aureus* has largely focussed on homologous recombination, which can affect genes sequenced in typing schemes, such as those used in MLST.

Analysis of MLST sequence data has shown that new variation in *S. aureus* is 15 times more likely to be the result of point mutation than recombination (Feil, Cooper et al. 2003). Compared to 48 other bacterial and archaeal species, *S. aureus* has the second lowest ratio, at 0.1, of r/m, the probability that a nucleotide change is due to the result of recombination rather than point mutation. To put the relative rate of *S. aureus* recombination into perspective, *Flavobacterium psychrophilum,* the species investigated with the highest recombination rate, has an estimated r/m of 63.6, and the median r/m in the study was 1.35 (Vos and Didelot 2009). However, these estimates were made using MLST data which may suffer from unknown biases to do with the length and location of fragments and may not be representative of the entire genome. Despite being a predominantly clonal organism recombination still provides a fundamental source of variation in *S. aureus*. An example of the role of recombination in *S. aureus* evolution is ST239, a very common ST associated with MRSA worldwide, which arose from a large recombination event between genomes belonging to ST30 and ST8 (Robinson and Enright 2004).

1.8.2 The role of natural selection in *S. aureus* evolution

Random genetic drift and natural selection both have an important role in creating and maintaining *S. aureus* diversity. It is interesting to distinguish between the two, in order identify genes that are under greater levels of selection. The d_N/d_S ratio can be used to indicate the likely forms of selection acting on a gene. Here d_N is the number of non-synonymous substitutions in the DNA sequence that alter amino acid coding and d_S the number of synonymous or silent

substitutions, which do not change amino acid coding. Values of d_N/d_S less than one are consistent with purifying selection, where most mutations changing amino acids reduce fitness. Values of d_N/d_S greater than one are a commonly recognised signature of positive, or directional selection (Yang and Bielawski 2000). Calculation of the d_N/d_S ratio from *S. aureus* MLST data gave a value of 0.77 within CCs and 0.12 for more distantly related samples. This suggests that in *S. aureus* non-synonymous mutations are selected against and removed over time, probably because they are deleterious (Feil, Cooper et al. 2003). However, these values may be underestimates as they are calculated on MLST genes, which are more likely to be under purifying selection. Despite this, the d_N/d_S ratio calculated from the *spa* gene in diverse isolates collected worldwide was similar, at 0.15 (Koreen, Ramaswamy et al. 2004).

1.8.3 Diversity and geographic spread

MLST data has shown that *S. aureus* clusters into 10 major CCs (Lindsay, Moore et al. 2006), all of which contain meticillin resistant isolates. Despite being present in each major lineage, MRSA has arisen recently from a subset of existing MSSA lineages on acquisition of SCC*mec,* so its molecular diversity remains much lower than that of MSSA (Enright, Day et al. 2000; Grundmann, Aanensen et al. 2010). MRSA also shows greater geographic structure than MSSA (Grundmann, Aanensen et al. 2010; Gray, Tatem et al. 2011), perhaps because the more recent origins of MRSA lineages means they have not had time disperse worldwide like MSSA has. A correlation with *S. aureus* spread and human migration has been demonstrated (Gray, Tatem et al. 2011), so with time, MRSA may become evenly distributed worldwide.

1.8.4 Core and accessory genome

Much work on the evolution of bacterial genomes uses gene sequences that are present in every isolate, however genomes are dynamic entities which gain and lose sequences at high rates through timescales relevant to the evolution of populations and species. In this context, “core” and “accessory” genomes are useful, if ill-defined, terms that use empirical findings about variation in bacterial gene content to make inferences about the genes required for survival and non-essential genes with specialised function (Lan and Reeves 2000). The 2.8 mega base (mb) *S. aureus* genome consists of approximately 75% core genes and 25% accessory genes (Lindsay and Holden 2006). Core genes encode functions vital for the existence of *S. aureus*, such as metabolism, DNA synthesis and replication, and determine properties characteristic of *S. aureus*.

The core genome has a lower observed mutation rate, perhaps because of purifying selection whereby many mutations are immediately selected against as they are detrimental to survival (Lan and Reeves 2000; Lindsay and Holden 2006; Malachowa and DeLeo 2010). The accessory genome includes genes with an array of non-essential functions including, notably, virulence and antibiotic resistance. Many of these genes are present on MGEs that can be transferred horizontally between *S. aureus* strains and even between species. The accessory genome explains more of the different phenotypic characteristics observed in *S. aureus* including adaptations required for *S. aureus* to survive in different ecological niches (Lindsay and Holden 2006; Malachowa and DeLeo 2010) than the core genome.

Aggregating the complete core and accessory genomes from many individuals of the same species would create a "species" or "pan"-genome (Lan and Reeves 2000). In species, such as *S. aureus,* with large gene pools and potential to recombine with other species, this may be constantly expanding (Medini, Donati et al. 2005).

1.8.5 Measuring genetic diversity

Like many other bacteria, *S. aureus* is a genetically diverse species which can also be divided into groups, such as CCs. When studying genetic diversity, questions of interest include: whether diversity is distributed mainly within or between groups, i.e. whether there is population structure; and whether measurement is at the nucleotide level or at a lower resolution, perhaps using results from molecular typing.

A simple descriptive method for quantifying genetic diversity within a group of samples is the population average pairwise difference (π_x). π_x is calculated from the mean number of differences between all pairs in a sample set, allowing comparison of π_x values between groups of different sizes. Additionally, Watterson proposed a statistic for estimating θ, which can also be used to quantify within group diversity, and similarly provides a number that can be compared between groups of different sizes. Watterson's θ simply counts the number of segregating sites then divides the total by a denominator to account for sample size. It is calculated using following formula (Watterson 1975):

$$\text{Watterson's } \theta = \frac{k}{(1/1 + 1/2 + 1/3 + 1/4 + \ldots + 1/(n-1))}$$

However, an underlying assumption of both π_x and Watterson's θ is that the sample used must be randomly selected from an evenly mixed population.

To investigate diversity between two or more groups π_{xy}, quantifies the average number of nucleotide differences between groups. π_{xy} is calculated by taking the mean number of differences between each pair of samples when each pair has one isolate from each group. After calculating diversity within and between groups it may then be desirable to quantify the proportion of diversity explained by each of these two measures. F_{ST} provides a measure of population structure: a value from zero to one, where one indicates that all the variation is explained by differences between the two groups.

Calculations of π and θ described above assume that groups are equivalent to populations. However, π can also be used to estimate diversity within groups, such as defined by CC. π therefore offers two advantages over Watterson's θ, firstly that it is appropriate to use for investigating group diversity rather than Watterson's θ which is only appropriate for measuring diversity of an entire population, and secondly, π has an equivalent for comparing diversity between groups, which Watterson's θ does not.

Finally, a suitable strategy for comparing diversity from different published studies, particularly when molecular typing techniques are used, is Simpson's Index (Limbago, Fosheim et al. 2009). Simpson's Index represents the probability that two randomly selected items from a sample set will be different, and accounts for both the number of types (for example STs) as well as their frequencies within a sample. Simpson's Index results in a value from zero to one, where a higher number indicates increased diversity, and is calculated using the following formula, where N represents the total number of individual organisms in the entire sample and n the number of individual organisms of each type:

$$\text{Simpson's Index} = \frac{\Sigma\, n(n-1)}{N(N-1)}$$

It is not possible to compare this formulation of Simpson's Index between sample sets of different sizes. However, population comparison is possible using a bootstrapping approach, whereby multiple subsets of the same size are taken from each sample, and Simpson's Index calculated

from each. This bootstrapping method provides the additional advantage that 95% confidence intervals (CI) can also be calculated (Grundmann, Hori et al. 2001), which if non-overlapping confirm two samples are different.

1.8.5.1 Phylogenetic trees

Phylogenetic trees are a method of graphically displaying evolutionary relationships within or between species. Many methods have been developed for drawing trees, which largely fall into two groups: distance and discrete methods. Distance methods include Neighbour Joining and Unweighted Pair Group Method with Arithmetic Mean (UPGMA), which estimate phylogenies by grouping together similar sequences. Discrete methods include Maximum Likelihood and Minimum Evolution, which identify the changes between nucleotide sequences and infer their position on the branches of the tree (Page and Holmes 1998).

When using phylogenetic trees to represent bacterial evolution, the extent of evolution by recombination, as represented by r/m (section 1.8.1), must be considered. If evolution is entirely clonal, as seen in *Leptospira interrogans* (Vos and Didelot 2009) then traditional methods for drawing phylogenetic trees (mentioned above) are likely to reconstruct the correct tree. However, even in these circumstances it is useful to assess the statistical support of the tree, either by using bootstrapping on one of the methods mentioned above (Page and Holmes 1998), or by using a Bayesian approach such as BEAST (Drummond and Rambaut 2007). On the other hand, if species have very high rates of recombination, such as *F. psychrophilum* (Vos and Didelot 2009), many points along the genome will have a different evolutionary history so it is best not to use phylogenetic trees at all. Instead, programs such as STRUCTURE (Pritchard, Stephens et al. 2000) account for different histories along a genome by assuming a number of populations, and that each individual comprises a mixture from these.

In the case of *S. aureus*, which is primarily clonal, with a low level of recombination (section 1.8.1), it is possible to draw phylogenetic trees; however the effect of recombination must be accounted for. Traditional tree drawing methods for *S. aureus* are likely to overestimate the evolution rate, as each recombination event may be considered multiple mutation events; and could result in star-like trees with overestimation of terminal branch lengths, which could incorrectly be inferred as population expansion.

The program ClonalFrame was designed to consider evolutionary relationships accounting for recombination and mutation (Didelot and Falush 2007) and is appropriate to use when drawing trees of species which are neither entirely clonal nor have very high recombination rates. ClonalFrame identifies regions in the genome which appear to have higher levels of mutation than expected and assigns them as regions of recombination, additionally it uses a Bayesian approach, so also provides a level of statistical support for the tree drawn. For species such as *S. aureus* with low rates of recombination, ClonalFrame is able to identify regions affected by recombination, whilst leaving enough data to accurately construct a phylogeny.

1.9 Whole-genome sequencing

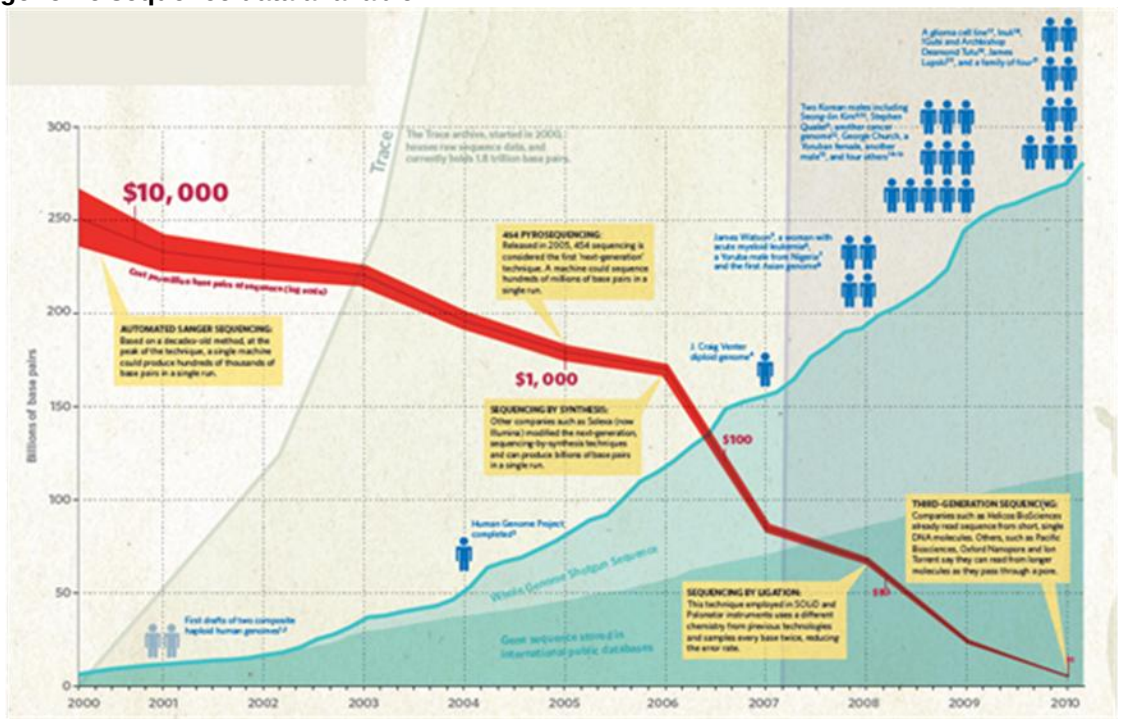
1.9.1 Next-generation sequencing technology

The publication of Sanger's method for DNA sequencing in 1977 provided the groundwork for over 30 years of sequencing using this technique (Sanger, Nicklen et al. 1977). Sanger sequencing initially generated sequences of up to 300 bases, and has since been modified to generate longer fragments thus allowing sequencing of entire genes and the development of techniques such as MLST. Shotgun sequencing, a process in which Sanger sequencing is used to sequence multiple overlapping, randomly generated, DNA fragments which are then aligned into longer, contiguous sequences, was used to create the first bacterial whole-genome sequence, that of *Haemophilus influenza* in 1995 (Fleischmann, Adams et al. 1995).

The advent of second-generation DNA sequencing, using methods other than Sanger sequencing, has allowed more DNA to be sequenced at a faster rate and lower cost, and resulted in an influx of genome sequences available for analysis (Figure 1. 4). In the past five years, multiple new technologies have been invented to produce complete genome data, the most successful of which are currently the 454 Genome Sequencer (Roche Applied Science; Basel, Switzerland), the Illumina platform (Illumina, San Diego, USA) and the SOLiD platform (Applied Biosystems; Foster City, CA, USA). These techniques all use cyclic-array sequencing, where DNA molecules are attached to separate locations on an array and sequenced using iterative cycles of enzymatic manipulation after which details of the sequence are collected using imaging of the array (Shendure and Ji 2008).

Figure 1.4: The sequence explosion (Nature 2010)

Red line is log of cost of one million bases of sequencing, blue line is amount of genome sequence data available.



Each next-generation platform uses different technology, resulting in different speed, costs, accuracy and fragment (read) length of DNA output. The Illumina platform was used for DNA sequencing in this project, and so will be explained in more detail.

Illumina sequencing begins with library preparation, where the DNA to be sequenced is sheared into approximately 200 base fragments which pass through two enzymatic steps to ensure they are of similar size and suitable for adaptor ligation. Adaptors are then attached to each end of the fragment, after which one adaptor from each fragment is bound to the surface of the “flow cell” or array. The unattached adaptor links to a complementary adaptor on the flow cell, forming a bridge. The fragment is amplified using bridge PCR, resulting in double stranded fragments. These then denature leaving single strands, for which the process can be repeated until clusters of PCR amplified fragments are formed on the flow cell.

Once the clusters have formed, sequencing begins by addition of four fluorescently labelled reversible chain terminators which bind to the end base of each fragment within a cluster. Lasers are used to photograph the flow cell, on which clusters light up one of four colours according to

the fluorescently labelled base in the cluster. Chain terminators are added in multiple steps and a photograph taken at each step, revealing the DNA sequence of each cluster in a series of images. The sequence of images from each cluster makes a read. Each read is one of multiple overlapping fragments of DNA that can then be assembled into a complete genome.

The first version of Illumina, GAI, introduced in 2006, used flow cells with eight lanes, and allowed sequencing of one sample per lane, with output in fragments of DNA or reads that were 36 bases long. After library preparation of three to four days a sequencing run for an approximately 2.8 mb genome such as *S. aureus* took one week, thus eight genomes could be sequenced in about 10 days. Since then many improvements have been made in the GAII and subsequently HiSeq platforms, lanes could be multiplexed by addition of six base nucleotide tags to the DNA fragments before sequencing, which enable identification of each sample. Illumina have now further improved the platform and by using 96-plex tags and 768 genomes can be sequenced in about 17 days.

1.9.2 Sequence assembly

Assembly of next-generation sequence reads is currently done using either of two basic approaches. Mapping-based assembly involves placing each read where it best matches a chosen reference sequence. For each position in the reference a number of reads will build up and a difference in the sample compared to the reference occurs at positions where all, or most of, the reads aligned differ from the reference (Figure 1.5). Base quality scores are a transformation of the error probability and can be calculated for each base in the reference, using information including the number of reads aligned to each position in the reference, their read qualities and the proportion of reads which disagree with the consensus. This gives a measure of how reliable the decision, or call, at each base is.

Figure 1.5: Alignment of reads to a reference visualised in Samtools tview

Green bases mark the reference. Blue bases highlight one individual read. Pink bases mark positions different from the reference. The line of pink bases signify a variant call where all positions are different from the reference.

Mapping to a reference has the disadvantage that sequence existing in the sample but not the reference will be lost. References must be carefully chosen, since the more similar the reference genome is to the samples being sequenced the better the quality of the alignment. However advantages of sequence assembly by mapping are that if many sequences are aligned to the same reference they can easily be aligned and compared.

De novo assembly involves piecing reads together without the aid of a reference sequence. With obvious advantages of not being reliant on genes present in a reference this may seem a preferable technique, however *de novo* assembly almost never produces a single complete genome, and output is often given in about 100 contigs, or genome fragments, which may be difficult to piece together. Results from *de novo* assembly are harder to align, so finding variants within a dataset is more complicated.

1.10 *S. aureus* genomics

The first whole-genome sequences of *S. aureus* were completed and annotated in 2001 using shotgun sequencing (Kuroda, Ohta et al. 2001). There are now 23 complete genomes available publicly at the National Centre for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/

as accessed 01/02/2011) as well as many more that have been sequenced using next-generation technology. Early publications of complete genome sequences included one or two genomes followed by detailed descriptions of genetic elements causing potential pathogenicity or antibiotic resistance (Kuroda, Ohta et al. 2001; Holden, Feil et al. 2004).

As next-generation sequencing developed, more genomes could be sequenced at once, allowing better inferences about population structure and evolution. Studies are now able to compare multiple isolates, previously indistinguishable by MLST, and make inferences about evolution on a much smaller scale. For example, whole-genome sequencing of 10 USA300 isolates revealed 578 unique single nucleotide polymorphisms (SNPs) in the genomes. Eight of the 10 isolates sequenced differed from each other by between only 11 to 48 SNPs, yet had different virulence in a mouse model, indicating the importance of whole-genome sequencing to identify small differences that affect virulence (Kennedy, Otto et al. 2008). Greater diversity was found within ST239, with 4310 SNPs in 63 isolates. Comparison of the two datasets using Watterson's Theta, a method for estimating population mutation rates whilst allowing for comparisons of populations of different sizes (Watterson 1975), showed that ST239 is a more diverse clone, perhaps reflecting its greater age and possibly wider sampling.

Whole-genome sequencing has also provided insight into *S. aureus* evolution rates. Estimates of evolution rates from the ST239 dataset were 3.3×10^{-6} mutations per site per year. This allowed calculation of the date for the most recent common ancestor for ST239 as mid to late 1960s, a date that corresponds with the emergence of MRSA into Europe (Harris, Feil et al. 2010). Similarly, sequencing of USA300 showed it was part of a very recent clonal expansion, in fitting with the recent increase in community-acquired disease seen in the USA (Kennedy, Otto et al. 2008).

Finally, whole-genome sequencing has increased understanding of *S. aureus* virulence. The first two complete sequences revealed almost all known *S. aureus* virulence factors, as well as 70 new candidates (Kuroda, Ohta et al. 2001). In 2005 when six *S. aureus* genomes had been completely sequenced, seven pathogenicity islands were identified in all six genomes, which were thought to carry approximately half of *S. aureus* virulence factors (Gill, Fouts et al. 2005).

Whole-genome sequencing is still in the early stages, with improvements constantly being made. As sequencing continues to reduce in cost and increase in quality, increasing numbers of studies will be performed more easily. Thus findings that have been made so far surrounding *S. aureus* diversity, evolution and virulence are only the tip of the iceberg in relation to what whole-genome sequencing will reveal about *S. aureus* in the future.

1.11 Thesis Outline

This thesis begins by using molecular techniques and epidemiology to examine potential definitions for community-acquired *S. aureus*, particularly MRSA, in Oxfordshire. This enabled estimation of CA-MRSA prevalence according to different definitions, and evaluation of definitions for identifying community-acquired isolates. Additionally, comparisons of different populations in Oxfordshire and Brighton allowed investigation of risk factors for MRSA bacteraemia and acquisition of MRSA in individuals without hospital contact.

Focus then moves to *S. aureus* in the healthy community, a much less studied group, where risk factors are not confounded by healthcare. To investigate this, a longitudinal carriage study was designed and individuals were enrolled and their *S. aureus* carriage followed over a period of up to 18 months. Demographic and epidemiological information was collected to determine why some individuals carry *S. aureus* whilst others do not, and of those who do carry, why some carry intermittently and others persistently. Molecular techniques were used to investigate the population structure of *S. aureus* in Oxfordshire and how strains carried in the nose changed over time.

Finally, whole-genome sequencing was used to examine two clonal groups that are responsible for the bulk of UK *S. aureus* disease and that in previous studies were each genetically uniform using other molecular typing techniques. Population structure across space and time was measured by sequencing isolates from two geographic locations and two time periods. Additionally, the high resolution of whole-genome sequencing allowed identification of an outbreak not previously possible with molecular typing.

2 Laboratory Materials and Methods

2.1 Sample collection

S. aureus isolates were obtained from various sources. Isolates for Chapter Three were obtained prospectively from the routine microbiology laboratories of Oxford Radcliffe Hospital (ORH) and Brighton and Sussex University Hospital National Health Service (NHS) Trust, between 2003 and 2006. They had all previously been identified as meticillin-sensitive or resistant *S. aureus.* Isolates were obtained as stocks on nutrient agar slopes at ambient temperature or from glycerol stocks (section 2.2.5). Samples for Chapter Five were obtained during the carriage study, whose design is outlined in Chapter Four. Samples were obtained as charcoal nasal swabs and testing for *S aureus* was carried out as in section 2.2.3. Finally, for Chapter Six, isolates from four collections were used. These consisted of invasive disease and healthy carriage isolates from Oxford, collected between 1997 and 1998 (Peacock, Moore et al. 2002); bacteraemia isolates collected from Brighton and Sussex University hospitals from 1999-2001 and 2006-2007 (Price, Baker et al. 2010); and a subset of the clinical and carriage isolates from ORH used in Chapter Three. All isolates used for Chapter Six were previously identified as meticillin-sensitive or resistant *S. aureus* and stored as glycerol stocks.

2.2 *S. aureus* identification and culture

2.2.1 Culture from nasal swabs

S. aureus was cultured from newly collected samples as follows: nasal swabs were first placed in 5% NaCl broth (E and O Laboratories, Bonnybridge, UK) overnight at 37 ^{o}C to inhibit growth of other bacteria. Following this, a 5 millimetre (mm) loopful of broth was inoculated onto SaSelect chromogenic agar (Bio-Rad, Limerick, Ireland) and incubated at 37 ^{o}C overnight. SaSelect chromogenic agar causes *S. aureus* to develop a characteristic pink colour as it contains a chromogen that detects phosphatase activity in *S. aureus.* Pink colonies were regarded as presumptive *S. aureus* and their identity was confirmed using DNase, catalase and Staphaurex tests, as described below. The majority of culture from nasal swabs and subsequent *S. aureus* identification was performed by Rowena Fung.

2.2.2 Culture of pre-identified *S. aureus*

Isolates on ambient slopes or glycerol stocks, previously confirmed to be *S. aureus,* were grown on Columbia Blood Agar (CBA) (E and O Laboratories, Bonnybridge, UK) at 37 ^{o}C overnight.

2.2.3 Identification of *S. aureus*

2.2.3.1 DNase test

The DNase test identifies degradation of DNA by the enzyme deoxyribonuclease produced by *S. aureus.* It was performed using DNase agar (E and O Laboratories, Bonnybridge, UK), which contains DNA within its medium. *S. aureus* was cultured onto the DNase agar and incubated overnight at 37 ^{o}C. Following this, 1 molar (M) hydrochloric acid (HCl) (VWR, Leicestershire, UK) was added to the agar, causing it to precipitate DNA, making the agar opaque. *S. aureus* was detected if bacterial colonies were surrounded by a clear region on the agar, produced due to the DNA being hydrolysed by the deoxyribonuclease enzyme.

2.2.3.2 Catalase test

The catalase enzyme present in *S. aureus* can be identified using the catalase test. Catalase decomposes hydrogen peroxide into water and oxygen, and is therefore detected by production of bubbles of oxygen when a small amount of *S. aureus* is added to hydrogen peroxide. To perform the test, *S. aureus* colonies were incubated overnight at 37 ^{o}C on CBA after which a colony of presumptive *S. aureus* was added to a capillary tube containing hydrogen peroxide. Bubbles in the capillary tube indicated the presence of *S. aureus*.

2.2.3.3 Staph Xtra Latex Test

The Staph Xtra Latex (Staphaurex) test detects presence of the clumping factor and protein A found on the *S. aureus* cell wall, using a rapid latex agglutination test (Pro-lab diagnostics, Cheshire, UK). To perform the Staphaurex test, *S. aureus* colonies were incubated overnight at 37 ^{o}C on CBA and emulsified into the Staph Xtra latex and the negative control latex reagents for 10-20 seconds (s). The Staph Xtra latex reagent contains latex particles coated with human immunoglobulin G and fibrinogen. Agglutination of the latex in the positive reagent upon addition of presumptive *S. aureus*, due to a protein-protein interaction between clumping factor and Protein A with immunoglobulin G and fibrinogen, indicated that the isolate was *S. aureus*.

Cultures were identified as *S. aureus* if they were positive for all three tests described above. If cultures were positive for two out of three tests, a further confirmatory tube coagulase test was performed.

2.2.3.4 Tube coagulase test

The tube coagulase test identifies the presence of the *S. aureus* enzyme coagulase. 0.5 millilitre (ml) rabbit plasma (Becton Dickinson, Oxford, UK) was added to a 2.5 ml bottle of peptone water (E and O Laboratories, Bonnybridge, UK), and one colony of presumptive *S. aureus,* previously incubated overnight at 37 ^{o}C on CBA, was emulsified into the tube. Tubes were incubated at 37 ^{o}C for a minimum of four hours. Agglutination of the rabbit plasma in the peptone water indicated presence of the coagulase protein.

If the culture was positive in the tube coagulase test and two out of the three initial tests, it was considered to be confirmed as *S. aureus*.

2.2.4 Determination of meticillin resistance

2.2.4.1 Antimicrobial susceptibility disc

S. aureus isolates were tested for β-lactam (including meticillin) resistance by growing overnight at 37 ^{o}C on CBA plus 5% sodium chloride (Oxoid, Basingstoke, UK) with an oxacillin antimicrobial susceptibility test disc (Becton Dickinson, Oxford, UK), containing 1 microgram (µg) oxacillin applied to the surface of the agar. Growth within a $\leq$14 mm diameter surrounding the disk indicated β-lactam and therefore meticillin resistance.

2.2.4.2 Etest

For isolates where the oxacillin susceptibility disc was unclear (i.e. close to 14 mm diameter) and for testing ciprofloxacin susceptibility, an Etest (Biomerieux, Marcy l'Etoile, France) was used. Two to three *S. aureus* colonies were emulsified into peptone water and spread evenly onto iso-sensitest agar (E and O Laboratories, Bonnybridge, UK) using a swab. The agar was dried for 10-15 minutes and the Etest strip applied to the agar. The culture was incubated at 37 ^{o}C overnight, following which the minimum inhibitory concentration (MIC) value could be read at the point where growth inhibition intersected the strip. MIC $\geq$ 4 indicated resistance for meticillin or ciprofloxacin; susceptibility was inferred with an MIC of $\leq$ 2 for meticillin and $\leq$ 1 for ciprofloxacin.

Intermediate sensitivity (values between the MICs for susceptible and resistant) was not seen in this study.

For more details on *S. aureus* growth and identification see Appendix 1 for *S. aureus* Standard Operating Procedures (SOP) for the carriage study.

2.2.5 Long term storage of bacterial cultures

Bacteria were stored as glycerol stocks, by emulsifying a fresh culture in 133 microlitres (µl) saline (E and O Laboratories, Bonnybridge, UK) containing 133 µl 45% glycerol (Sigma, Dorset, UK) to constitute a 200 µl 15% glycerol stock solution. Approximately 50 different colonies were emulsified into a 600 µl solution, which was used to make a master and working stock for each isolate, and were stored at -80 ^{o}C

2.3 DNA extraction and PCR

2.3.1 DNA extraction methods

DNA was extracted using three different methods, depending on the study.

2.3.1.1 QIAGEN DNeasy tissue kit

DNA used in Chapter Three, was extracted using the Qiagen DNeasy tissue kit (Qiagen, Crawley, UK). A single colony was emulsified in 5% NaCl (E and O Laboratories, Bonnybridge, UK) overnight at 37 ^{o}C and then extracted according to manufacturer's instructions. Briefly, cells were lysed using proteinase K, and DNA was purified by binding to a silica-membrane, washing with ethanol and eluting in 10 millimolar (mM) Tris·Chloride and 0.5 mM Ethylenediaminetetraacetic acid (EDTA).

2.3.1.2 Crude chromosomal DNA extract

Crude chromosomal DNA extracts were prepared from isolates collected during the carriage study. A mixed sweep of the culture on SaSelect agar was added to a mixture of 10 µl 0.85% saline and 50 µl Tris EDTA (TE) buffer (Sigma, Dorset, UK). This was heated at 99.9 ^{o}C for 10 minutes, then clarified by centrifugation at of 16.1 x gravity (G) using an Eppendorf 5810R centrifuge (Eppendorf, Cambridge, UK) for two minutes. The supernatant, containing the chromosomal DNA (50 µl) was removed and stored at -20 ^{o}C.

2.3.1.3 Homogenisation followed by silica binding

Illumina sequencing of isolates for Chapter Six required higher concentrations of pure DNA. Cultures incubated on CBA overnight at 37 ^{o}C were lysed using the FastPrep homogeniser (MP Biomedicals, Illkirch, France) with lysing Matrix B, according to the protocol for gram positive bacteria, with two 40 s homogenisations at a setting of 6.0 on the FastPrep instrument. DNA was purified from the cell lysate by binding to a silica gel suspension, washing with ethanol and elution into DNase free water. DNA concentration was determined using a NanoDrop spectrophotometer (Labtech International, East Sussex, UK), and verified using the PicoGreen dsDNA Assay Kit (Invitrogen, Paisley, UK). 1.25 micro grams (µg) of DNA was required for each sequencing reaction at a concentration of 25 nanograms (ng)/µl.

2.3.2 PCR

2.3.2.1 *mecA* and Panton Valentine Leukocidin (PVL) confirmation

A multiplex PCR was used to test *S. aureus* isolates for *lukS*/*F-PV* (genes encoding PVL) and *mecA*, together with 16s rRNA (used as a positive PCR control) (McClure, Conly et al. 2006). Isolates were tested for the *mecA* gene to confirm they were MRSA. PVL is commonly associated with community-acquired *S. aureus* disease so genes encoding PVL were tested for when investigating CA-MRSA in Chapter Three. The published method for the multiplex PCR was modified to use a final volume of 25 µl containing 1 µl template DNA, 2.5 µl 10 x Qiagen PCR Buffer (Qiagen, Crawley, UK), 0.5 mM magnesium chloride ($MgCl_2$) (Qiagen, Crawley, UK), 5 micromolar (µM) each deoxyribonucleotide (dNTP) (Invitrogen, Paisley, UK), 0.125 µl Hotstart Taq DNA polymerase (Qiagen, Crawley, UK) and 0.01 µM, 0.04 µM and 0.02 µM of forward and reverse primers for the three gene targets; 16S, PVL and *mecA* respectively (Table 2.1) (Operon Scientific, Cologne, Germany). Modifications to the published conditions were made to ensure optimal assay performance on the thermal cycler available: a PTC-200 Peltier Thermal Cycler (MJ Research, Boston, Mass). Thermocycling conditions were set at 15 minutes (min) of denaturation at 95 ^{o}C, followed by 35 cycles of 30 s denaturation at 95 ^{o}C, 30 s primer annealing at 52 ^{o}C and 1 min extension at 72 ^{o}C, with a final extension for 5 min at 72 ^{o}C. A clinical MRSA isolate from ORH, isolate 212 (16S rRNA+, lukS/FPV-, mecA+) and V7 (16S rRNA+, lukS/FPV+, mecA-) (a gift of Dr A. Kearns, Heath Protection Agency) were used as positive PCR controls.

PCR amplicons were visualised using 1% agarose (Severn Biotech ltd, Worcester, UK) gel electrophoresis prepared by heating with 1 x Tris Borate EDTA (TBE) buffer (Sigma, Dorset, UK). Ethidium bromide (Sigma, Dorset, UK) was carefully added to the slightly cooled liquid gel at a final concentration of 0.5 µl/ml, which was set at room temperature. 5 µl of PCR-amplified DNA was mixed with 0.2 µl gel loading buffer (Sigma, Dorset, UK) and 0.8 µl distilled water and added to the wells of the agarose gel. A 100 bp ladder (Invitrogen, Paisley, UK) was used as a molecular weight marker. Electrophoresis was carried out at 120 volts with 1 x TBE as the electrophoresis buffer (Sigma, Dorset, UK).

Table 2.1: Primers used for Multiplex PCR

Gene	Primer	Sequence	Amplicon size (bp)	Reference
16s rRNA	Staph765F Staph 750R	AACTCTGTTATTAGGGAAGAACA CCACCTTCCTCCGGTTTGTCACC	756	(McClure, Conly et al. 2006)
lukS/F-PV	Luk-PV-1 Luk- PV-2	ATCATTAGGTAAAATGTCTGGACATGATCCA GCATCAAGTGTATTGGATAGCAAAAGC	433	(McClure, Conly et al. 2006)
mecA	MecA1 MecA2	GTAGAAATGACTGAACGTCCGATAA CCAATTCCACATTGTTTCGGTCTAA	310	(McClure, Conly et al. 2006)

2.4 Molecular Typing

2.4.1 MLST

2.4.1.1 PCR

MLST was performed essentially as described (Enright, Day et al. 2000) using improved oligonucleotide primers designed for the present study. Improved primer pairs were designed using the Oligonucleotide Properties Calculator (Kibbe 2007). Primers were designed to be between 18-25 bases long with GC content between 35-65%, annealing temperature between 55-65 ^{o}C ending with the bases C and G. Primers were checked for self-self annealing and entered into BLAST (http://blast.ncbi.nlm.nih.gov/Blast) to confirm that they were specific for *S. aureus*. DNA amplification was performed in a PTC-200 Peltier Thermal Cycler in a final volume of 25 µl containing 2.5 µl 10 x Qiagen PCR buffer, 5 mM Taq DNA polymerase, 0.02 µM forward and reverse primer, 5 µM each dNTP and 1 µl template DNA. Primers are described in Table 2.2. For the genes *arcC*, *aroE*, *pta*, *tpi* and *yqiL* thermocycling conditions were set at 2 min of denaturation at 94 ^{o}C, followed by 35 cycles of 30 s denaturation at 94 ^{o}C, 30 s annealing at 50 ^{o}C, and 1 min

extension at 72 ^{o}C, with a final extension at 72 ^{o}C for 5 min. For the genes *glpF* and *gmk* conditions were identical except annealing was carried out at 55 ^{o}C for 30 s.

Amplification products were visualised using a 1% agarose gel. Amplified DNA was purified by precipitation using 60 µl 20% Polyethylene glycol/2.5 molar (M) NaCl (Sigma, Dorset, UK) for 30 min before centrifuging using an Eppendorf centrifuge 5415D (Eppendorf, Cambridge, UK) at 2750 G at 4 ^{o}C for 60 min. The resultant pellet was washed twice using 150 µl 70% ethanol (VWR, Leicestershire, UK). After ambient drying, the DNA was re-suspended in between 5 and 50 µl of molecular grade water (Invitrogen, Paisley, UK). The re-suspension volume was dependent on the intensity of the PCR product observed in the agarose gel.

Oligonucleotide sequencing of the PCR product was performed in a primer extension reaction containing a final volume of 10 µl comprising 1.87 µl 5 x Buffer (consisting of 400 mM tris HCl and 10 mM $MgCl_2$ (both from Sigma, Dorset, UK)), 0.25 µl ABI Prism Bigdye Terminator (Applied Biosystems, Warrington, UK); 4 µl oligonucleotide primer (0.67 µM) and 1.5 µl template DNA (the product from initial DNA amplification). The reaction conditions were 30 cycles of 96 ^{o}C for 10 s, 50 ^{o}C for 5 s and 60 ^{o}C for 2 min, again using a PTC-200 Peltier Thermal Cycler.

The extension products were diluted with 10 µl molecular grade water and precipitated with 50 µl 100% ethanol (VWR, Leicestershire, UK) and 2 µl 3 M sodium acetate (pH 5.2) (Sigma, Dorset, UK). This was incubated at room temperature for 45 min before centrifugation at 2750 G at 4 ^{o}C for 60 min followed by one wash with 150 µl 70% ethanol. After air drying the reaction products were analysed using the ABI 3730xl DNA instrument at the Sequencing Facility, Zoology Department, University of Oxford.

Table 2.2: Primers used for MLST

Gene	Primer	Sequence	Amplicon size (bp)	Reference
Carbamate kinase (*arcC*)	arcc F arcc R	TTGATTCACCAGCGCGTATTGTC AGGTATCTGCTTCAATCAGCG	456	(Enright, Day et al. 2000)
Shikimate dehydrogenase (*aroE*)	aroe F aroe R	GCAGTTATCGGAAATCCTATTTCAC CTCATTAAAGTATTGGGAGAAAGATGC	456	Designed for this study
Glycerol kinase (*glpF*)	glpf F glpf R	TTTGGTGGTGGCGTTTGTG CCTAATAAACCACCGGCAATTGG	465	Designed for this study
Guanylate kinase (*gmk_*)	gmk F gmk R	GTTAATCGTTTTATCAGGACCATC GTTCATCAATTTCACGCGCTC	429	Designed for this study
Phosphate acetyltransferase (*pta_*)	pta F pta R	GTTAAAATCGTATTACCTGAAGG GCTTCTTGAACTTTTGTCACGTCG	474	pta F: (Enright, Day et al. 2000) pta R: This study
Triosephosphate isomerase (*tpi_*)	tpi F tpi R	TCGAAGATAATGGTGCGTTCACAG ACCATGTTCGCTTTCGCGGTTCG	402	Designed for this study
Acetyl coenzyme A acetyltransferase (*yqiL*)	yqiL F yqiL R	CGAGAGTCGTATTAGCAGCAGC GGTTCACCTTTACGTTGAGGAATCG	516	Designed for this study

2.4.1.2 MLST sequence determination

Forward and reverse chromatograms of each gene required for MLST were assembled using STARS (Sequence Typing Analysis and Retrieval System) available at http://sara.molbiol.ox.ac.uk/userweb/mchan/stars/adminguide.htm, which is an alternative interface to Staden (Staden 1996). Resulting DNA consensus sequences were edited where necessary, trimmed using a Perl script written by Dr David Wyllie, and entered into the *S. aureus* MLST database (http://saureus.mlst.net/). Here MLST allele numbers were assigned for each locus, and the resulting seven digit allelic profile was submitted to obtain a sequence type for each isolate. Novel alleles were confirmed by repeat sequencing and allocated a new allele number by the database curator. eBurst (Feil, Li et al. 2004) was used to group STs with six out of seven MLST alleles in common with at least one other member of the group into clonal complexes.

2.4.2 *spa* typing

2.4.2.1 PCR

spa typing was performed as described previously (Harmsen, Claus et al. 2003), using the published oligonucleotide primers where possible. *spa* typing produces amplicon sizes ranging

from one to 23 repeats of the 21-30 bp repeat unit. Due to the small size of some of the *spa* type amplicons, modified primer pairs were used to increase amplicon size where necessary. Primers used for *spa* typing are described in table 2.3. DNA amplification and sequencing reaction conditions were as for the MLST gene *arcC* (section 2.4.1.1). Amplification was performed either using a PTC-200 Peltier Thermal Cycler (MJ Research, Boston, Mass) or DNA was sent to the Oxford Centre for Gene Function where it was amplified using the ABI 3730 DNA Analyzer (Life Technologies, USA). The majority of *spa* typing for the carriage study in Chapter Five was performed by Dr Antonina Votintseva.

Table 2.3: Primers used for *spa* typing

Primer	Sequence	Reference
1095F	AGACGATCCTTCGGTGAGC	(Harmsen, Claus et al. 2003)
1517R	GCTTTTGCAATGTCATTTACTG	
spa2for	AAMYGAAGAACAACGTAACGGC	Designed for this study
spa2rev	TAATAACGCTGCACCTAASG	

2.4.2.2 Assignment of *spa* types

Forward and reverse chromatograms for the *spa* gene were assembled using Ridom StaphType (Harmsen, Claus et al. 2003), which identifies the number and order of *spa* repeats in the sequence, assigns them a *spa* type and then submits this to the *spa* database (http://www.ridom.de/spaserver/). BURP clustering (Mellmann, Weniger et al. 2007) was used to group *spa* types into clonal complexes by assigning a cost to each repeat unit duplication and excision as well individual point mutations, and then clustering isolates related by a cost of four or less.

3 MRSA in the Community?

3.1 Introduction

The first report of MRSA acquired in the community was in 1980 in a group of intravenous drug users (Diederen and Kluytmans 2006). Since then CA-MRSA has increased in prevalence, especially in the USA where it is now common, largely presenting as skin and soft tissue infections (Davis, Perri et al. 2007). In the UK, CA-MRSA is uncommon and was recently shown to represent <1% of MRSA (Elston and Barlow 2009). Whilst the majority of UK MRSA remains predominantly healthcare associated, CA-MRSA has been reported to be increasing (Verena Schneider-Lindner 2007). Despite this, the true level of MRSA in the UK community, its genetic profile and how it presents all remain unknown.

Definitions of CA-MRSA are highly varied and group into three main forms, often dependent on the information available. Firstly, definitions can be based on suspected origins of MRSA, which cannot be known for certain. Origin based definitions are often used when further epidemiological information is unavailable and have advantages in their simplicity. An example of an origin-based definition of community acquisition, used in this chapter, is the isolation of MRSA from a patient on admission to hospital, which selects individuals with a pre-existing infection who therefore have not acquired MRSA on their current hospital admission. However, such infections could have been acquired during recent previous hospital contact.

Secondly, empirical definitions are based upon detailed epidemiological information about the individual and classify CA-MRSA based on various healthcare risk factors. These often vary depending on the information available, but include the CDC definition for CA-MRSA, defined as: MRSA diagnosed in outpatient or within 48 hours after admission to hospital; no medical history of MRSA infection or colonisation; no history in the previous year of: hospitalisation; admission to a nursing home, skilled nursing facility or hospice, dialysis or surgery; and no permanent indwelling catheters or medical devices that pass through the skin into the body (Naimi, LeDell et al. 2003).

Finally, MRSA can be defined as community-acquired based on microbiological information, independent of the source of isolation (Millar, Loughrey et al. 2007). Community-acquired

samples have been reported commonly to have STs distinct from common hospital associated strains; carry resistance to fewer antibiotics; and harbour the PVL toxin (Grundmann, Aires-de-Sousa et al. 2006; Millar, Loughrey et al. 2007). An example of a microbiological definition is ciprofloxacin sensitivity, which acts as a simple microbiological marker that may enrich for CA-MRSA as it is commonly resistant to fewer antibiotics (Grundmann, Aires-de-Sousa et al. 2006). Microbiological criteria may be used independently from epidemiological data to identify possible CA-MRSA, or in conjunction with origin-based or empirical definitions. If a strain fits an origin-based or empirical definition of CA-MRSA and has a genetic background indicative of CA-MRSA it is highly likely to be CA-MRSA (Millar, Loughrey et al. 2007).

In this chapter, three case-control studies, each using one of the three forms of CA-MRSA definition described above, will investigate CA-MRSA in Oxfordshire. This will enable determination of whether there is CA-MRSA in Oxfordshire and its distinguishing characteristics, it will also enable evaluation of definitions of CA-MRSA.

3.2 Hypothesis

The study was designed to investigate the following hypothesis:

> There is community-acquired MRSA in Oxfordshire which has a distinguishable epidemiology and population structure

3.2.1 Importance of the hypothesis

CA-MRSA in the UK is under-investigated, with studies focusing on specific outbreaks (Orendi 2006; Adedeji, Weller et al. 2007; Atkinson, Paul et al. 2009). At the time of designing this study its prevalence was unknown, although was thought to be uncommon. In the USA, where CA-MRSA is a problem, it is the most common identifiable cause of skin and soft-tissue infections (Groom, Wolsey et al. 2001; Hussain, Boyle-Vavra et al. 2001; Moran, Krishnadasan et al. 2006; Chambers and Deleo 2009). Accurate knowledge of CA-MRSA prevalence in the UK will establish whether CA-MRSA is likely to cause problems such as seen in the USA and will inform empiric antibiotic treatment policies.

In order to investigate the hypothesis in detail, three definitions for CA-MRSA were used, increasing the likelihood of identifying CA-MRSA and its characteristics, whilst enabling

evaluation of methods for CA-MRSA identification. It is important to reliably identify CA-MRSA so hospitals can identify specific populations at higher risk and change their policies accordingly, for example by testing more people on admission to ensure they do not transmit MRSA into the hospital or altering antibiotic regimens. In addition, some evidence suggests that CA-MRSA may be more virulent and more easily transmitted than HA-MRSA (Deleo, Otto et al. 2010). CA-MRSA could therefore act as a reservoir for advantageous strains, which if they were to acquire additional antibiotic resistance determinants may be able to outcompete or even replace current epidemic HA-MRSA strains leading to new epidemics. It is therefore important to understand the likelihood of this risk.

3.2.2 Novelty of the hypothesis

Previous studies investigating CA-MRSA in the UK have focused on specific outbreaks (Orendi 2006; Adedeji, Weller et al. 2007; Atkinson, Paul et al. 2009), or used only molecular definitions (Health Protection Agency 2005; Otter and French 2008). Therefore, although presence and characteristics of CA-MRSA have been investigated before in the UK, the hypothesis has wider implications, since this study combines three CA-MRSA definitions and uses a larger sample set than ever before in the UK, making it more in depth than previous studies.

3.3 Methods

In order to identify cases of CA-MRSA in Oxfordshire three case-control studies were designed, each with different criteria to identify possible community acquisition. The selection criteria were: (i) individuals with MRSA bacteraemia on admission to hospital; (ii) individuals with MRSA who had not been in hospital in the year prior to isolation; and (iii) individuals with ciprofloxacin sensitive MRSA. To compare the characteristics of these groups with potential CA-MRSA, each set of cases was compared to a control group of the same size. Control groups were as follows group (i) to individuals with MSSA bacteraemia on admission to hospital; group (ii) to individuals with hospital contact in the year prior to baseline; and group (iii) to individuals with ciprofloxacin resistant MRSA.

Cases from study one were selected using an origin based definition of CA-MRSA. This definition was chosen because 5% of individuals in the Oxfordshire area are estimated to have been hospitalised in the previous year (Wyllie, Walker et al. 2007), therefore more stringent definitions

may rule out individuals with hospital exposure in the previous year who acquired MRSA elsewhere. Isolates were taken on admission to hospital to ensure patients had not acquired MRSA during their current hospital stay, and 48 hours was chosen to define MRSA on admission in fitting with the CDC definition (section 3.1). Additionally, individuals with MRSA bacteraemia were selected because individuals with infection on admission are a significant group representing 27% of the total MRSA bacteraemia reported in England and Wales 2006-2007 (Health Protection Agency 2007) and were important to investigate, since bacteraemia is a significant cause of mortality (Cosgrove, Sakoulas et al. 2003).

Following this, cases from study two were selected because research from the USA has shown that CA-MRSA may be less likely to present as bacteraemia (Davis, Perri et al. 2007). Therefore all types of MRSA sample, including screening samples, taken from individuals with no *S. aureus* clinical presentation, were used. However, to increase the stringency of case selection, individuals were excluded if they had any hospital exposure in the year prior to MRSA isolation, in fitting with epidemiological definitions of CA-MRSA, such as the CDC definition (section 3.1).

Finally, cases from study three were selected based upon ciprofloxacin sensitivity, a microbiological definition of CA-MRSA. A microbiological definition was chosen to compliment the first two case-control studies because it may identify a different set of CA-MRSA strains that were never transmitted in hospital and may have spread under different selective forces from nosocomial MRSA.

For each of these studies, only samples sent into hospital for testing were included, therefore mild infections, which may not have been reported to the ORH laboratory would be missed. However, experience suggests that CA-MRSA infections are often severe and require treatment (Adedeji, Weller et al. 2007; Davis, Perri et al. 2007) so a clinically significant subset of MRSA would probably not go unnoticed.

To fully investigate epidemiological and microbiological risk factors socio-demographic and healthcare related information was collected for individuals in each group to determine whether they had CA-MRSA and investigate its characteristics. The genetic profiles of nosocomial and community MRSA were investigated using MLST and *spa* typing to characterise the population

structure and compare between cases and controls. Detailed methods for each of the case-control studies are outlined below.

3.3.1 Case-control study one: Admission bacteraemia

3.3.1.1 Sample selection

Individuals were selected to be MRSA admission bacteraemia cases or MSSA admission bacteraemia controls if they had MRSA or MSSA bacteraemia isolated on admission to the general medical service of Oxford Radcliffe Hospital Trust between April 2003 and September 2006. 'On admission' was defined as having *S. aureus* isolated within 48 hours of admission to hospital, consistent with the CDC definition of CA-MRSA (Naimi, LeDell et al. 2003). Bacteraemia included all samples isolated from blood.

Cases were selected by searching ORH electronic records for any individuals with *S. aureus* bacteraemia who had been in hospital for less than 48 hours. All MRSA isolates within the time period meeting the inclusion criteria were included. Two sets of MSSA controls were chosen. The first was unmatched and consisted of an equal number of MSSA admission bacteraemia samples selected from within the study period. The second was age-matched, chosen because the first set of controls revealed that age was a potential confounder for other factors. Here, every MRSA case was matched to an MSSA control with an age difference of fewer than six years, also selected from within the study period.

3.3.1.2 Epidemiological information collected

For each participant, ORH electronic records were used to identify date of most recent in-patient and out-patient appointment in ORH, the total number of in-patient and out-patient appointments in the previous year in ORH, and whether the participant had MRSA isolated previously.

A structured questionnaire was completed from case notes for each MRSA case and MSSA control (both unmatched and age-matched). The questionnaire was designed to identify demographic information; risk factors for *S. aureus* infection (O'Sullivan and Keane 2000; Lodise, McKinnon et al. 2003; Naimi, LeDell et al. 2003; Salgado, Farr et al. 2003; Tacconelli, Venkataraman et al. 2004; Pan, Diep et al. 2005); and information about clinical presentation, medications and participant outcomes; thus assisting in identification of community-acquired

samples and comparison of epidemiology between cases and controls. Details included in the questionnaire are listed in table 3.1 and a copy of the questionnaire in Appendix 2.

3.3.1.3 Microbiology

Isolates were obtained from ambient slopes that had previously been determined as meticillin-resistant or meticillin-sensitive *S. aureus*. Multiplex PCR and MLST were carried out on all isolates according to the Methods. *spa* typing, the method of choice in later studies, had not yet been introduced in our laboratory at the time of this study.

3.3.1.4 Ethical approval

The study was conducted as an audit; therefore ethical approval was not required.

3.3.2 Case-control study two: MRSA in individuals without previous hospital contact

3.3.2.1 Sample selection

All individuals with MRSA isolated from a clinical or screening sample between 13 February-22 April 2007 and 24 July-11 November 2007 who had not been an in-patient or out-patient in the previous year according to ORH electronic records (other than in the previous 24 hours) were considered as possible community cases. Structured questionnaires identifying risk factors for MRSA acquisition were sent to the GPs of potential participants to ascertain whether they had had any acute hospital contact in the year prior to MRSA isolation. Those that were confirmed not to have had any in-patient or out-patient exposure in any acute hospital in the year prior to MRSA isolation, from electronic records and the GP questionnaire were community cases. For each community case a hospital-exposed control was identified as an individual with MRSA isolated on the same or the consecutive day who had been an in-patient or out-patient in ORH in the year prior to MRSA isolation.

3.3.2.2 Epidemiological information collected

ORH electronic records were used to identify potential community cases who had MRSA isolated and had not been an in-patient or out-patient in ORH in the year prior to MRSA isolation. Basic details about hospital-exposed controls including demographics and information about recent hospital exposure were also collected.

The questionnaire sent to GPs of potential community cases was designed to identify hospital exposure in the previous year that was not documented in electronic records, as well as other healthcare risk factors recorded in GP records. A copy of the questionnaire is available in Appendix 3. GPs were sent the questionnaire by fax; if there was no response within one week they were telephoned as a reminder, if there was still no response after a further week, they were followed up through the practice manager. The questionnaire collected information on dates, durations and where applicable, locations, of acute and non-acute hospital in-patient stays; hospital out-patient visits; nursing or care home residence; regular care from practice nurses; indwelling catheters or invasive procedures; and antibiotics in the year prior to MRSA isolation. The questionnaire was not sent to hospital-exposed controls as they were already known to have hospital exposure.

Individuals identified through the questionnaire to have no risk factors, or only antibiotic risk factors, were investigated further through GP records to fully confirm that they had no healthcare risk factors.

3.3.2.3 Microbiology

Isolates were obtained from ambient slopes that had previously been determined as meticillin-resistant or meticillin-sensitive *S. aureus*. Multiplex PCR, MLST and *spa* typing were carried out on all isolates according to the Methods.

3.3.2.4 Ethical approval

This, and the following study (section 3.3.3) were covered by Statutory Instrument Regulations 2002 No. 1438, section (iii) “Communicable disease and other risks to public health (Health Service Control of Patient Information)” of section 60 of the Health and Social Care Act (www.opsi.gov.uk/si/si2002/20021438.htm) in line with Health Protection Agency policy; therefore MREC (Multi-Centre Research Ethics Committee) approval was not required.

3.3.3 Case-control study three: Ciprofloxacin Sensitive (CipS) MRSA

3.3.3.1 Sample selection

CipS MRSA cases for this study were initially selected to include all clinical and screening samples over three time periods: 11 June 2006-31 January 2007, 13 February-22 April 2007 and

24 July-11 November 2007, which were sensitive to ciprofloxacin according to ORH electronic records. CipS cases were matched to an equal number of ciprofloxacin resistant (CipR) controls from the same time periods. To better represent the potential diversity of CipS cases an additional set of CipS isolates from Brighton and Sussex University hospitals collected between January-October 2007 was included, without controls.

3.3.3.2 Epidemiological information collected

ORH electronic records were used to collect information on basic demographics and previous ORH hospital exposure of Oxford CipS cases and CipR controls. Similar information about the Brighton CipS strains was sent from Brighton and Sussex University hospitals.

3.3.3.3 Microbiology

Isolates from Oxford were obtained from ambient slopes and from Brighton from glycerol stocks. Both had previously been determined as meticillin-resistant or meticillin-sensitive *S. aureus*. Multiplex PCR, MLST and *spa* typing were carried out on all isolates as according to the Methods. Ciprofloxacin sensitivity was confirmed using Etest, as in the Methods.

3.3.4 Statistical methods

For all three case studies statistical analysis was performed using SPSS version 14.0 (SPSS Inc., USA). To describe continuous data, medians and inter-quartile ranges were used as they are not skewed by extreme variables. Univariate analysis was performed using Fisher's Exact Test to compare categorical variables and Mann-Whitney rank sum tests for continuous variables. These are both non-parametric methods and were used because the data did not fit a fixed probability distribution. Multivariate analyses were performed, to rule out confounding variables and consider the effects of multiple variables together, using backwards stepwise logistic regression. This removed variables not considered significant one after another and then recalculated the significance of the remaining model to ensure a final model that included all relevant independent variables. Time-to-event outcomes were censored at three years and analysed using Kaplan-Meier plots and log rank tests. *P* values of <0.05 were considered statistically significant.

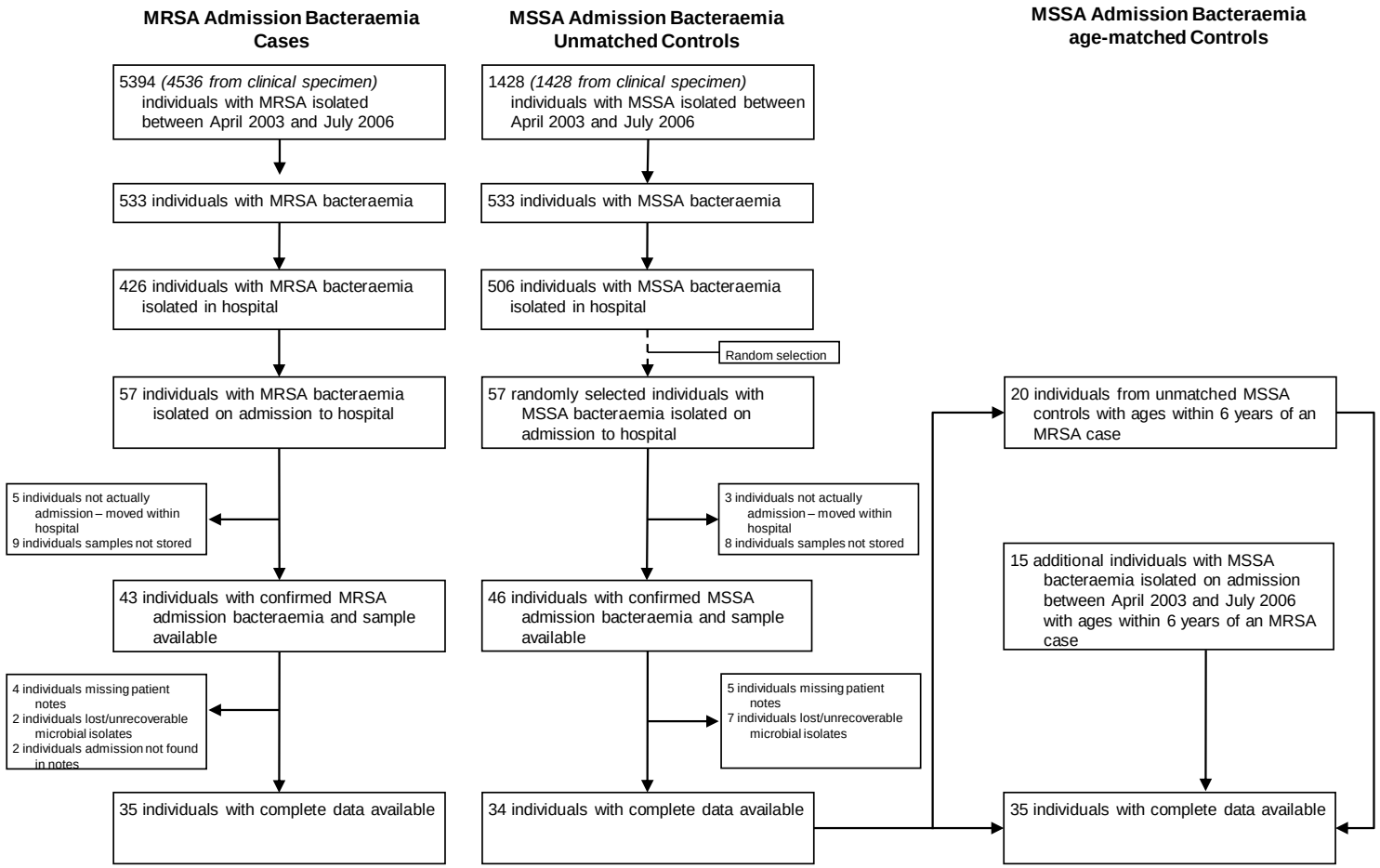
3.4 Results

3.4.1 Case-control study one: Admission bacteraemia

3.4.1.1 Recruitment

Fifty-seven MRSA admission bacteraemia cases were identified within the study period (April 2003-September 2006), from 5394 MRSA samples sent to the John Radcliffe Hospital microbiology laboratory for routine testing (samples from the John Radcliffe, three surrounding hospitals and GP practices in Oxfordshire), of which 553 were bacteraemia and 426 were bacteraemia samples from individuals in hospital. Fifty-seven individuals with MSSA admission bacteraemia were randomly selected as controls (Figure 3.1). On further investigation, 14 MRSA cases and 11 MSSA unmatched controls did not meet inclusion criteria, either because they were not truly admission bacteraemia, having moved within the hospital (five MRSA cases, three MSSA controls) or because they were admitted to the Horton General hospital, where isolates were not stored (nine MRSA cases, eight MSSA controls). From the 43 MRSA cases and 46 MSSA unmatched controls left, it was possible to obtain all the information required for the questionnaire for 35 MRSA cases and 34 unmatched MSSA controls. Each field in the questionnaire was missing from a maximum of two participants, apart from body temperature, which was missing from three cases. Thirty-five age-matched MSSA controls were also identified, of which 20 were already part of the unmatched MSSA control dataset and 15 were identified from the same period to ensure all controls matched an MRSA case within six years of age (Figure 3.1).

Figure 3.1: Sample selection for Admission bacteraemia study



3.4.1.2 Epidemiological characteristics

To investigate the characteristics of MRSA admission bacteraemia cases, the epidemiology of the cases was compared to that of the MSSA admission bacteraemia controls. MRSA admission bacteraemia cases had a median age of 81 years, compared to 70 years for MSSA admission bacteraemia controls (P<0.001). Because MRSA cases were older, age was a potential confounder when comparing these two groups. Therefore analysis was carried out comparing cases to age-matched MSSA admission bacteraemia controls who had a median age of 80 years, not significantly different from MRSA admission bacteraemia cases (P=0.34).

Univariate analyses comparing MRSA admission bacteraemia cases and age-matched MSSA admission bacteraemia controls revealed that MRSA cases were more likely to be male (P=0.007) (Table 3.1). They were also more likely to have previous hospital exposure, by measures that included: fewer days since last in hospital (more than half of MRSA cases had been admitted in the last two months (median days since last admitted for MRSA cases 51 versus 236 for MSSA controls, P<0.001)), hospital admission in the last year, (97% of MRSA cases had been admitted to hospital in the previous year (P<0.001)), more in-patient and out-patient appointments in the year prior to *S. aureus* bacteraemia (P=0.001 and P=0.09 respectively), and more days spent in hospital in the year prior to *S. aureus* bacteraemia (P<0.001). MRSA cases also tended to be more likely to have had MRSA isolated previously (P=0.05) (Table 3.1).

Due to the small numbers of individuals with each healthcare risk factor, analysis of healthcare contact was grouped into three measures. The first was whether individuals fit the CDC definition of CA-MRSA (section 3.1) (Naimi, LeDell et al. 2003). The second, whether individuals had a plausible healthcare portal of entry, which was any of: invasive surgery or intervention in the year prior to *S. aureus* bacteraemia, vascular access at the time of *S. aureus* bacteraemia, urinary catheter at the time of *S. aureus* bacteraemia or any previous renal dialysis or admission into intensive care. Finally, individuals were given a co-morbidity score from zero to six, with one point for each of the following concurrent conditions: diabetes, heart failure, oral steroids, chronic obstructive pulmonary disease (COPD), vascular disease and chronic renal failure.

Results from univariate analysis of these groups showed that no MRSA admission bacteraemia cases fit the CDC definition for community-acquired disease compared to 13/35 (37%) of age-matched MSSA controls (*P*<0.001). MRSA cases were more likely to have a healthcare portal of entry (*P*=0.003) and tended to have a higher co-morbidity score than age-matched controls (*P*=0.005) (Table 3.1).

Measures of clinical presentation did not differ between MRSA cases and MSSA controls, who had very similar levels of illness severity. However, all participants had abnormalities in at least one clinical measure and over 80% of participants in each group had been judged clinically to have an infection on the day of presentation (Table 3.1).

Table 3.1: Characteristics of MRSA admission bacteraemia cases compared to MSSA unmatched and age-matched controls

Characteristic	**MRSA (N=35) N (%) or median (interquartile range (IQR))†**	**MSSA unmatched controls (N=34) N (%) or median (IQR)†**	***P* value versus MRSA cases**	**MSSA age-matched controls (N=35) N (%) or median (IQR)†**	***P* value versus MRSA cases**
Demographics:					
Sex (male)	26 (74%)	20 (88%)	0.21	14 (39%)	0.007
Age†	81 (75;87)	70 (57;81)	<0.001	80 (70;84)	0.34
Admitted from:					
Nursing home	5 (14%)	0 (0%)	0.08	2 (6%)	0.70
Residential home	4 (11%)	3 (9%)		4 (11%)	
Community hospital	4 (11%)	2 (6%)		3 (9%)	
Own home	22 (63%)	28 (82%)		26 (74%)	
Hostel	0 (0%)	1 (3%)		0 (0%)	
Previous hospital exposure:					
Admitted to hospital in the past year	34 (97%)	16 (47%)	<0.001	19 (54%)	<0.001
Admitted to hospital in the past three years	34 (97%)	21 (62%)	<0.001	21 (60%)	<0.001
Days since last in hospital†	51 (10;118)	591 (14;1361)	<0.001	236 (19;1361)	<0.001
Number of in-patient appointments in last year†	1 (1;2)	0 (0;1)	<0.001	0 (0;1)	0.001
Number of out-patient appointments in last year†	1 (0;3)	0 (0;1)	0.05	0 (0;1)	0.09

Characteristic	MRSA (N=35) N (%) or median (interquartile range (IQR))†	MSSA unmatched controls (N=34) N (%) or median (IQR)†	*P* value versus MRSA cases	MSSA age-matched controls (N=35) N (%) or median (IQR)†	*P* value versus MRSA cases
Number of days in hospital in the past year†	17.4 (1.9;33)	0 (0;3)	<0.001	0 (0;3)	<0.001
Co-morbidities:					
Healthcare portal of entry	25 (71%)	12 (35%)	0.004	15 (43%)	0.03
invasive surgery	12 (34%)	4 (12%)	0.04	8 (23%)	0.43
invasive intervention	7 (21%)	6 (18%)	1.00	4 (11%)	0.34
vascular access	0 (0%)	0 (0%)		0 (0%)	
urinary catheter	11 (31%)	2 (6%)	0.01	4 (11%)	0.08
renal dialysis	1 (3%)	0 (0%)	1.00	0 (0%)	1.00
admission into intensive care unit	5 (14%)	1 (3%)	0.20	0 (0%)	0.05
Ulcers, eczema or psoriasis	8 (23%)	11 (32%)	0.43	7 (20%)	1.00
Intra-venous drug use	0 (0%)	3 (9%)	0.11	0 (0%)	1.00
Co-morbidity score†	1 (1;2)	0 (0;1)	0.002	0 (0;2)	0.05
diabetes	6 (17%)	4 (12%)	0.73	4 (11%)	0.73
heart failure	6 (17%)	2 (6%)	0.26	3 (9%)	0.48
oral steroids	2 (6%)	2 (6%)	1.00	2 (6%)	1.00
vascular disease	20 (57%)	9 (26%)	0.02	13 (37%)	0.15
COPD	4 (11%)	2 (6%)	0.67	3 (9%)	1.00
chronic renal failure	5 (14%)	1 (3%)	0.20	4 (11%)	1.00
CDC community-acquired:					
Fit CDC definition	0 (0%)	16 (47%)	<0.001	13 (37%)	<0.001
MRSA previously isolated:					
Previous MRSA	9 (26%)	1 (3%)	0.01	2 (6%)	0.05
Clinical presentation:					
Patient judged clinically to have an infection on the day of presentation	31 (89%)	28 (82%)	0.51	29 (83%)	0.73
Temperature ^{o}C†	37.2 (36.3;38.5)	37.8 (36.1;38.5)	0.80	37.4 (36.0;38.3)	0.62
Pulse†	88 (71;105)	109 (88;122)	0.003	97 (81;115)	0.09
Systolic blood pressure (BP)†	126 (105;143)	122 (100;140)	0.70	130 (100; 147)	0.74
Diastolic BP†	61 (54;75)	67 (57;82)	0.12	65 (57;85)	0.16
White cell count 10^9/litre (L) †	15.3 (10.9;17.8)	14.4 (8.8;20.8)	0.85	16.4 (10.3;21.4)	0.82
C-reactive protein milligram (mg)/L†	177 (92 253)	164 (80;267)	0.97	210 (143;285)	0.23
Neutrophils 10^9/L	12.6 (8.7;16.3)	12.5 (7.1;18.5)	0.93	13.9 (7.6;18.9)	0.80
Lymphocytes 10^9/L†	0.9 (0.6;1.4)	0.8 (0.6;1.5)	0.92	0.8 (0.6;1.2)	0.62
Platelets 10^9/L†	252 (176;336)	207 (88;274)	0.09	235 (129;357)	0.53

Characteristic	MRSA (N=35) N (%) or median (interquartile range (IQR))†	MSSA unmatched controls (N=34) N (%) or median (IQR)†	*P* value versus MRSA cases	MSSA age-matched controls (N=35) N (%) or median (IQR)†	*P* value versus MRSA cases
Haemoglobin g/decilitre†	12.3 (10.5;13.4)	12.5 (10.2;13.4)	0.93	12.1 (10.0;13.0)	0.49

† data represented by median (IQR)

All factors significant at *P*<0.10 (excluding the CDC definition as no MRSA cases fit this; and whether the patient had been in hospital in the previous year as this was nearly collinear with whether they had been in hospital in the previous three years) in the univariate analysis were placed in a backwards stepwise multivariate model, with whether an individual had MRSA or MSSA admission bacteraemia as the dependent variable. Variables left in the model were: male sex, with an odds ratio (OR) of 3.34, indicating that a male is 3.34 times as likely to have MRSA bacteraemia than a female; total number of days the participant spent in hospital in the previous year (OR=1.07 per day); and number of months between *S. aureus* isolation and previous hospital exposure (OR=0.94 per month) (Table 3.2).

Table 3.2: Significant characteristics differing between MRSA cases and MSSA controls using a multivariate backwards stepwise regression

Characteristic	Lower confidence interval (LCI)	Upper confidence interval (UCI)	Odds Ratio
Male	0.91	12.27	3.34
Days spent in hospital in previous year	1.07	1.00	1.13
Months since last in hospital	0.94	0.89	1.00

3.4.1.3 Microbiological characteristics

To determine whether MRSA admission bacteraemia cases had a distinguishable population structure, STs of MRSA admission bacteraemia cases were compared to those of MSSA admission bacteraemia controls. All but one MRSA admission bacteraemia case was ST22 (n=27) or ST36 (n=7), both associated with hospital acquired MRSA in the UK. The additional isolate was ST256, an

uncommon ST not known to be healthcare-associated. The individual with this ST was in their late 60s, admitted from their own home, with healthcare risk factors of hospital admission six months previously and co-morbidities of COPD and heart-failure. In contrast, unmatched MSSA admission bacteraemia controls had much more varied STs. The most common were ST15 and ST30 (both n=6) with 20 different STs in total. Healthcare-associated ST22 and ST36 were found in one and zero MSSA samples respectively, and the majority of samples were not associated with healthcare (Table 3.3). All MRSA admission bacteraemia cases and no MSSA controls had the *mecA* gene and no isolates had the PVL gene, both identified by multiplex PCR.

Table 3.3: Sequence type of admission bacteraemia MRSA cases and MSSA unmatched controls

Clonal Complex	ST	Number MRSA cases	Number unmatched MSSA controls
45	All STs	1	3
	45		2
	924 (Single Locus Variant (SLV) from ST45)		1
	256 (SLV from ST44)	1	
30	All STs	7	7
	30		6
	36 (SLV from ST30)	7	
	34 (SLV from ST30)		1
8	All STs		2
	8		1
	630 (SLV from ST8)		1
Singletons			
	1		1
	2		1
	5		2
	10		1
	12		1
	15		6
	20		1
	22	27	1
	25		1
	78		1
	182		1
	188		3
	464		1
	925		1
Total		35	34

3.4.1.4 Identification of CA-MRSA

Investigation of cases with admission MRSA bacteraemia did not reveal any individuals with CA-MRSA. No MRSA admission bacteraemia cases fit the epidemiology based CDC definition for CA-MRSA and 97% of cases with MRSA had epidemic hospital associated strains. This could represent the true absence of CA-MRSA in the time period, however, could also be because origin based definitions of CA-MRSA are not sufficiently rigorous, or because CA-MRSA largely presents as skin and soft tissue infection and not bacteraemia (Elston and Barlow 2009). However, the study did identify a specific population admitted to hospital with MRSA bacteraemia, who are predominantly

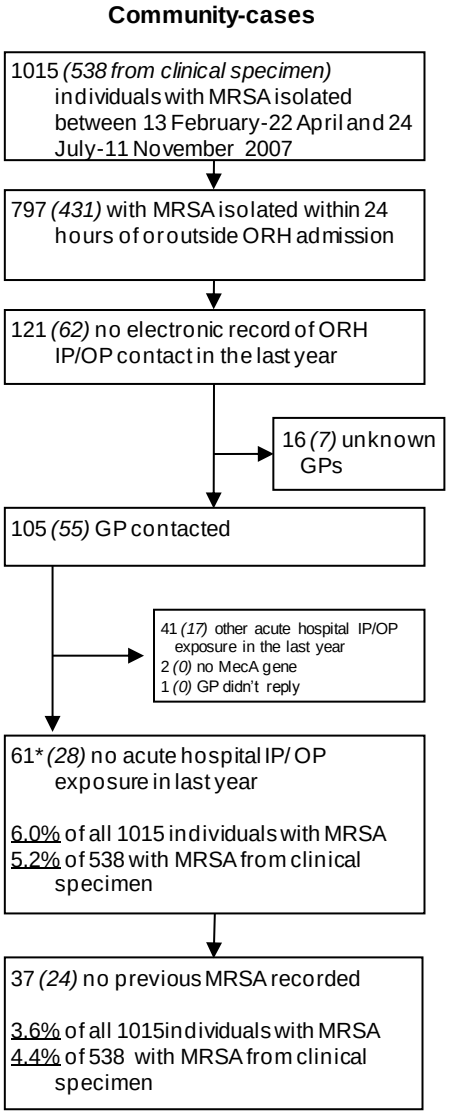
elderly with high levels of previous hospital exposure, but mostly live in their own homes prior to admission.

3.4.2 Case-control study two: MRSA in individuals without previous hospital contact

3.4.2.1 Recruitment

Electronic records identified 1015 individuals with MRSA (538 clinical and 477 screening samples) isolated in the two time periods (13 February-22 April 2007 and 24 July-11 November 2007) (Figure 3.2). Of these, 797 had MRSA isolated outside of hospital or within 24 hours of admission, and 121 had no electronic record of in-patient or out-patient appointment in ORH. Questionnaires were sent to the GPs of the 105 participants with known GPs, of which all but one replied. Of these, 61 were confirmed to have no acute hospital in-patient or out-patient exposure in the year prior to MRSA isolation, so were termed community cases (Figure 3.2). Thus, 97% of individuals had been in hospital in the year prior to MRSA isolation. 6.0% (61/1015) of individuals with MRSA samples (clinical or screening) sent into ORH had not had ORH hospital contact in the previous year, and 5.2% (28/538) were clinical samples. Sixty-one hospital-exposed controls were also identified.

Figure 3.2: Sample selection for community cases



3.4.2.2 Epidemiological characteristics

Univariate analysis was initially performed to identify any distinguishable characteristics of community cases compared to hospital-exposed controls. This showed that community cases were last discharged from hospital a median of 34 months ago, as opposed to hospital-exposed controls who, although selected to have hospital exposure in the previous year, had usually experienced very recent hospital exposure with a median of 12 days between discharge and MRSA isolation (P<0.001) (Table 3.4). Forty-seven (77%) samples from community cases were sent from GPs as opposed to only 11 (18%) hospital-exposed controls (P<0.001), however a substantial proportion of community

case samples were sent from acute hospital on admission (n=6, 10%), all of which were screening samples. Univariate analyses of other details available from electronic records showed that community cases were significantly older (*P*=0.001) and more likely to be male (*P*=0.002), compared to hospital-exposed controls (Table 3.4).

Table 3.4: Characteristics of community cases versus hospital-exposed controls

Four community cases and four hospital-exposed controls had missing susceptibilities to antimicrobials and were classed as sensitive

	Community case (N=61) n (%) or median (IQR)†	**Hospital-exposed control (N=61) n (%) or median (IQR)†**	***P* value**
Age (years)	82 (64;90)	72 (55;81)	0·001
Sex (male)†	19 (31%)	37 (61%)	0·002
Previous ORH in-patient or out-patient visit recorded in:			
Last three months	0 (0%)	56 (92%)	
Last year	0 (0%)	61 (100%)	
Last three years	35 (57%)	61 (100%)	<0·001
Ever	51 (84%)	61 (100%)	0·001
Days since last in ORH as in-patient or out-patient†	1020 (626;1835)	12 (0;44)	<0·001
Hours previously spent in ORH as in-patient†	41 (0;1162)	698 (47;1716)	0·005
MRSA previously isolated at ORH (any sample)	24 (38%)	19 (31%)	0·45
Antibiotic resistance			
Ciprofloxacin	56/57 (98%)	59/59 (100%)	0·49
Erythromycin	34/60 (57%)	35/59 (69%)	0·85
Fusidic acid	2/60(3%)	1/58 (2%)	1·00
Gentamicin	0/60 (0%)	2/59 (3%)	0·24
Mupirocin	2/60 (3%)	3/57 (5%)	0·67
Netilmicin	0/60 (0%)	0/58 (0%)	1·00
Oxacillin	57/60 (95%)	59/59 (100%)	0·24
Penicillin	60/60 (100%)	59/59 (100%)	1·00
Rifampicin	3/60 (5%)	0/58 (0%)	0·24
Tetracycline	2/60 (3%)	6/59 (10%)	0·16
Vancomycin	0/59 (0%)	0/60 (0%)	1·00
Number of resistant antibiotics out of 10 (excluding oxacillin)†	3 (2;3)	3 (2;3)	0·74
Sample sent from:			
GP	47 (77%)	11 (18%)	<0·001
Acute hospital <24 hours	6 (10%)	17 (29%)	
Acute hospital >24 hours	0 (0%)	27 (44%)	
Non-acute/community hospital	6 (10%)	6 (10%)	
Nursing home	1 (2%)	0 (0%)	
Sample type:			
Screening sample	23 (38%)	34 (56%)	0·11
Surface culture	33 (54%)	20 (33%)	
Surface culture and screen	2 (3%)	3 (5%)	
Other clinical sample	3 (5%)	2 (3%)	

† data represented by median (IQR)

To identify the effects of the independent variables together on whether an individual was a community case or hospital-exposed control, the data were analysed using backward stepwise

logistic regression. Independent variables entered in the regression were age, sex, days spent in ORH in the year prior to MRSA isolation, previous MRSA isolation, the sum of antibiotic resistances and where the sample was sent from. This confirmed that community cases were significantly more likely to be older (OR=1.03 per year increase in age), more likely to be female (OR=2.42) and spent fewer days in ORH in the year prior to MRSA isolation (OR=0.80 per day reduction in total stay), in addition community cases were more likely to have had MRSA previously isolated (OR=7.60) than hospital-exposed controls (Table 3.5).

Table 3.5 Significant risk factors for community cases versus hospital-exposed controls from backward stepwise regression

Characteristic	LCI	UCI	OR
Age	1.01	1.05	1.03
Female	1.06	5.54	2.42
Log days spent in hospital in previous year	0.56	0.39	0.80
MRSA previously isolated	2.96	1.15	7.60

3.4.2.3 Detailed epidemiology of community cases

Further investigation to determine whether community cases had a distinguishable epidemiology was performed using information from the questionnaire sent to GPs. The most common risk factor identified from the questionnaire was antimicrobials in the year prior to MRSA isolation (n=46, 75%), followed by regular care in the previous year (n=34, 56%) of which skin care (for example wound dressings) (n=23, 38%) was the most common form (Table 3.6). Cases had a median of 2/5 possible risk factors (excluding previous MRSA isolation), with regular care and antibiotics the most common combination of risk factors.

Eighteen (30%) community cases fit the CDC definition for CA-MRSA. Of these only three had no other risk factors and thirteen (56%) had received regular care and antibiotics or antibiotics only (ten and three respectively). Community cases not fitting the CDC definition for CA-MRSA were divided into those who had MRSA previously and those with no previous MRSA (n=24 and n=19 respectively). Univariate comparisons of the three groups (CDC CA-MRSA, non-CDC CA-MRSA with previous MRSA and non-CDC CA-MRSA no previous MRSA) showed that apart from factors in the

CDC definition for CA-MRSA, individuals with MRSA who fit the CDC definition for community acquisition did not significantly differ from those who did not.

Table 3.6: Characteristics of community cases from the GP questionnaire

	All community cases (N=61) n (%)	**CDC CA-MRSA‡ community cases (N=18) n(%) or median (IQR)†**	**Non-CDC CA-MRSA community - cases with previous MRSA (N=24) n (%) or median (IQR)†**	**Non-CDC CA-MRSA community - cases with no previous MRSA (N=19) n (%) or median (IQR)†**	***P* value across three groups**
Age†	82 (64;90)	79 (52;87)	82 (64;90)	84 (64;95)	0·42
Meets CDC definition	18 (30%)	18 (100%)	0 (0%)	0 (0%)	
MRSA previously isolated	24 (38%)	0 (0%)	24 (100%)	0 (0%)	
Previous in-patient or out-patient visit in ORH ever	51 (84%)	13 (72%)	23 (96%)	15 (79%)	0·08
Risk factors from GP questionnaire:					
Non-acute hospital in-patient in the last year	7 (11%)	0 (0%)	2 (8%)	5 (26%)	0·05
Non-acute hospital out-patient in the last year	13 (21%)	0 (0%)	3 (13%)	10 (53%)	<0·0001
Residence in a nursing or residential home in the last year	20 (33%)	0 (0%)	10 (42%)	10 (53%)	<0·0001
Regular care in the last year	34 (56%)	10 (56%)	13 (54%)	11 (58%)	1·00
Skin care	23	4	10	9	
Nursing care	5	2	2	1	
Regular medication	1	1	0	0	
Catheter	1	0	1	0	
Unspecified care	4	3	0	1	
Catheterised or invasive procedure in the last year	8 (13%)	0 (0%)	7 (29%)	1 (5%)	0·01
Antibiotics in the last year	46 (75%)	10 (56%)	20 (83%)	16 (84%)	0·09
Total number of risk factors†	2 (1;3)	1 (1;2)	2 (1;3)	3 (2;3)	0·0001
Risk factors from GP questionnaire					
none	4 (7%)	3 (17%)	1 (4%)	0 (0%)	0·02
Antibiotics only	8 (13%)	5 (28%)	3 (13%)	0 (0%)	
2 or more, or 1 (other than antibiotics) only	49 (80%)	10 (56%)	20 (83%)	19 (100%)	

† data represented by median (IQR)

Eight community cases with either no risk factors (n=3) or previous antibiotics as their only risk factors (n=5) from the GP questionnaire were further investigated using GP records. Four of these were found to have risk factors that had not been reported in the questionnaire: nursing home residence (n=1); regular GP or nursing care (n=2); or had been a non-acute out-patient in the previous year (n=1). Of the other four, two worked as nurses and one had acute hospital contact 13 months prior to MRSA isolation, leaving one individual with no healthcare risk factors at all.

3.4.2.4 Microbiological characteristics

MLST and *spa* typing were performed to investigate whether community cases had a distinguishable population structure. Fifty-eight cases and 58 controls were successfully typed using MLST (three isolates in each group were lost). The most common STs were UK healthcare associated ST22 (79% and 81% of community cases and hospital-exposed controls respectively) and ST36 (12% and 10% respectively). There was no significant difference in the distribution of CC amongst cases and controls (P=0.27) (Table 3.7). Community cases also included two individuals with STs in CC22 that differed by either one SNP or one single base pair deletion from ST22 (both of which were new STs first described in this study) and three individuals with STs in CC5, founded by ST5: a globally distributed lineage often found associated with healthcare (Witte, Cuny et al. 2008). All isolates had the *mecA* gene and none had the PVL gene.

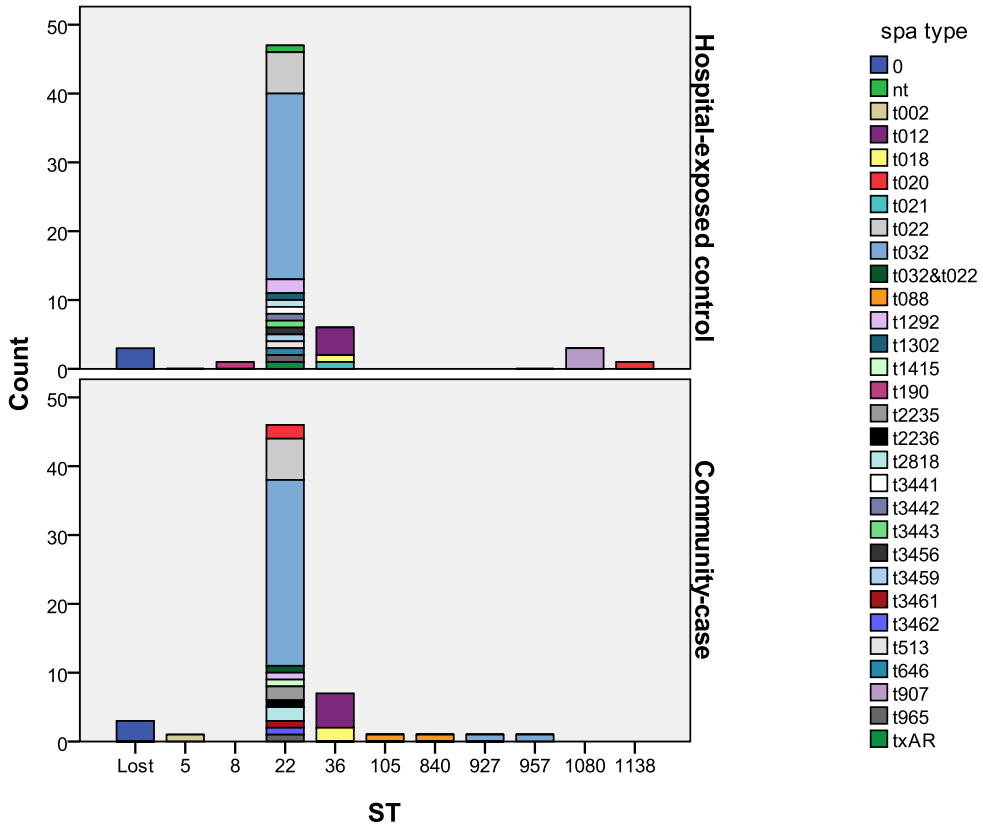
Table 3.7: MLST results of community cases and hospital-exposed controls

Strain marked with * is also a ciprofloxacin sensitive case.

	Sequence Type	Community case (N=61)		Hospital-exposed control (N=61)
		CDC CA-MRSA (N=18)	Non CDC CA-MRSA (N=43)	
Lost			3	3
Total known isolates		***18 (100%)***	***40 (93%)***	***58 (100%)***
Clonal complex (CC)5	***All STs***	***2 (11%)***	***1 (2%)***	***0 (0%)***
	5		1	
	105 (SLV from ST5)	1		
	840 (SLV from ST 5)	1*		
CC8	***All STs***	***0 (0%)***	***0 (0%)***	***1 (2%)***
	8			1
CC22	***All STs***	***13 (72%)***	***35 (81%)***	***51 (88%)***
	22	13	33	47
	927 (SLV from ST22)		1	
	957 (Single base deletion from ST22)		1	
	1080 (SLV from ST22)			3
	1138 (SLV from ST22)			1
CC30	***All STs***	***3 (17%)***	***4 (9%)***	***6 (10%)***
	36 (SLV from ST22)	3	4	6

spa typing was able to further discriminate isolates within ST22 and ST30. These were separated into 11 and two types respectively for community cases and 14 and three types for hospital-exposed controls (Figure 3.3). In general, *spa* typing was consistent with MLST results: individual STs all fell within the same *spa* CC except for ST22 for which two community cases and two hospital-exposed controls were of unknown *spa* CC and one community case was a singleton and could not be grouped with any other *spa* type.

Figure 3.3: STs separated by *spa* type for community cases and hospital-exposed controls



3.4.2.5 Epidemiology by ST

Epidemiological and microbiological information was then combined to identify any community cases that fit both empirical and microbiological definitions for CA-MRSA. The three community cases with MRSA belonging to the globally-distributed CC5, with ST5, ST105 and ST840, had two (non-acute out-patient and three courses of antibiotics in previous year), one (skin care) and zero risk factors respectively. The individual with ST840 and no risk factors according to the questionnaire was also the individual who when further investigated using GP records was found to have no healthcare risk factors at all. This individual also had ciprofloxacin sensitive MRSA and was therefore included in case-control study three. The other individual whose hospital exposure 13 months previously was only discovered by further examination of GP records had MRSA belonging to ST22. Individuals with ST22, but *spa* types in unusual *spa* CCs all had regular care as a risk factor and two had also had a course of antibiotics in the previous year.

3.4.2.6 Identification of CA-MRSA

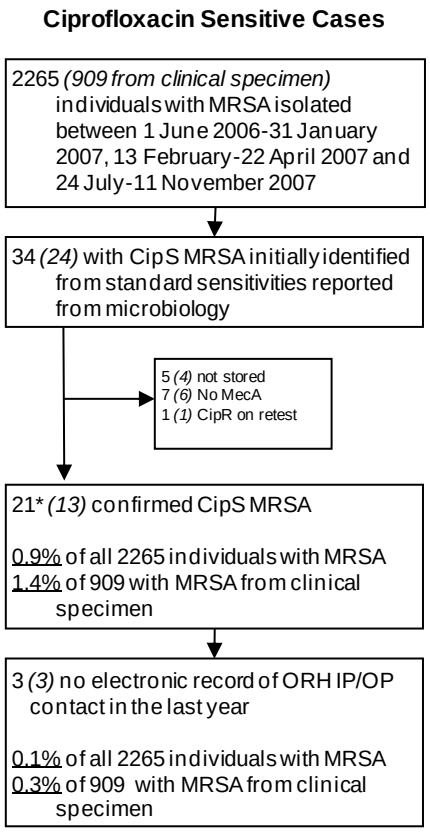
Investigation of community cases with no hospital exposure in the year prior to MRSA isolation reliably identified one individual with CA-MRSA out of 1015 investigated. This individual fit the CDC definition for CA-MRSA, had no other healthcare risk factors identified from investigation of GP records, and had an ST not associated with hospital. This suggests that epidemiological definitions are useful to identify CA-MRSA, however, detailed investigation is required. As an artefact of the epidemiological definition used in this study, 15 individuals with no healthcare risk factors, as defined by the CDC definition for CA-MRSA, but hospital associated STs were also identified. Isolates from these individuals may represent MRSA acquired during other non-hospital healthcare such as regular care in the community, as identified by the questionnaire, or prolonged carriage of MRSA acquired in hospital over a year before. Although they cannot be classified as CA-MRSA, these isolates are clearly distinguishable from most healthcare-acquired samples and could be termed “feral” or community-associated.

3.4.3 Case-control study three: CipS MRSA

3.4.3.1 Recruitment

Two-thousand two hundred and sixty-five individuals in Oxford had MRSA isolated over the three time periods (11 June 2006-31 January 2007, 13 February-22 April 2007 and 24 July-11 November 2007) (Figure 3.4), of which 34 were initially identified as CipS using standard sensitivities reported from ORH microbiology. Of these, five samples were not stored, seven had no *mecA* gene and one was not ciprofloxacin resistant on re-testing, leaving 21 confirmed CipS MRSA cases, each with a CipR control. Therefore 0.9% (21/2265) of individuals with MRSA (clinical or screening samples) sent into ORH have strains sensitive to ciprofloxacin, of which 1.4% (13/909) were clinical samples. Eighteen CipS samples were identified from Brighton and Sussex University hospitals.

Figure 3.4: Sample selection for Oxford CipS cases



3.4.3.2 Epidemiological characteristics

Univariate analyses were initially used to identify any distinguishable epidemiological characteristics of CipS cases compared to CipR controls. Results showed that compared to CipR controls Oxford CipS cases were: significantly younger (median age 37 compared with 74 years, $P<0.001$); had been in hospital longer ago (median of 28 days since last in ORH compared with six days, $P=0.03$); had spent fewer hours as an in-patient in ORH in the previous year ($P<0.005$); and yielded samples that were resistant to fewer antibiotics ($P<0.001$). Additionally Oxford CipS cases tended to be less likely to have had MRSA isolated previously (4/21 cases having previous MRSA versus 11/21 controls, $P=0.05$) (Table 3.8). Oxford CipS cases and CipR controls were not distributed differently between surface cultures, screens or other sample types (Table 3.8).

CipS samples from Oxford and Brighton were not significantly different from each other when compared using the predictors in table 3.8, except that Brighton CipS samples were more likely to

have been in hospital before (*P*=0.025) and Brighton samples did not include screening samples. The difference in hospital exposure could be explained by a different hospital record system in Brighton.

CipS samples from Brighton were not added to analysis comparing characteristics of CipS and CipR strains, since they were not directly matched with CipR samples from the same population as the Oxford CipS samples. However, addition of Brighton CipS into a comparison between CipS and CipR strains did not change the significance of any risk factors, apart from previously noted measures of previous hospital exposure and sample type, and additionally previous MRSA isolation, which became significantly less likely in CipS cases (*P*=0.008).

Table 3.8: Characteristics of CipS cases versus CipR controls

Seven CipS and six CipR antibiotic susceptibilities missing and recorded as sensitive.

	Oxford CipS case (N=21) n (%) or median (IQR)†	**Oxford CipR control (N=21) n (%) or median (IQR)†**	***P* value CipS case versus CipR control (Oxford)**	**Brighton CipS (N=18) n (%) or median (IQR)†**
Age (years)†	37 (13;67)	74 (65;82)	<0·001	56 (21;80)
Sex (male)	13 (62%)	9 (43%)	0·35	10 (56%)
Previous ORH or Brighton in-patient or out-patient visit recorded in:				
Last three months	14 (67%)	19 (90%)	0·13	9 (50%)
Last year	18 (86%)	21 (100%)	0·23	10 (56%)
Last three years	18 (86%)	21 (100%)	0·23	10 (56%)
Ever	18 (86%)	21 (100%)	0·23	10 (56%)
Days since last in ORH as in-patient or out-patient†	28 (3;105)	6 (0;28)	0·03	84 (0;-)
Hours previously spent in ORH as in-patient†	79 (0;318)	722 (392;1346)	<0·005	<0·005
MRSA previously isolated at ORH (any sample)	4 (19%)	11 (52%)	0·05	3 (17%)
Antibiotic resistance				
Erythromycin	6/21 (29%)	18/21 (86%)	<0·001	10/18 (56%)
Fusidic acid	7/15 (7%)	0/15 (0%)	0·006	9/18 (50%)
Gentamicin	0/18 (0%)	0/21 (0%)	1·00	0/18 (0%)
Mupirocin	3/21 (15%)	0/20 (0%)	0.23	1/18 (6%)
Oxacillin				
Penicillin	20/21 (95%)	21/21 (100%)	1·00	18/18 (100%)
Rifampicin	1/21 (5%)	1/20 (5%)	1·00	1/18 (6%)
Tetracycline	1/21 (5%)	0/21 (0%)	1·00	3/18 (17%)
Vancomycin	0/21(0%)	0/21 (0%)	1·00	0/18 (0%)
Number of resistant antibiotics out of nine (excluding oxacillin)†	2 (1;2·5)	3 (3;3)	<0·001	2 (1;3)
Sample sent from:				
GP	6 (29%)	8 (38%)	0·47	7 (39%)
Acute hospital <24 hours	6 (29%)	4 (19%)		4 (22%)
Acute hospital >24 hours	4 (19%)	7 (33%)		4 (22%)
Non-acute or community hospital	5 (24%)	2 (9%)		3 (17%)
Nursing home	0 (0%)	0 (0%)		0 (0%)
Sample type:				
Screening sample	8 (38%)	6 (29%)	0·30	0 (0%)
Surface culture	12 (57%)	10 (48%)		11 (61%)
Surface culture and screen	0 (0%)	0 (0%)		0 (0%)
Other clinical sample	1 (5%)	5 (24%)		7 (39%)

† data represented by median (IQR)

The Oxford-only data was analysed using a multivariate regression with backwards stepwise elimination, comparing CipS cases and CipR controls with age, days between previous hospital exposure and MRSA isolation, hours previously spent as an ORH in-patient and previous MRSA

isolation as independent variables. This confirmed that Oxford CipS cases and CipR controls differed significantly in age (OR=0.95), however, no other variables remained significant once age had been taken into account.

3.4.3.3 Microbiological characteristics

To identify whether CipS cases had a distinguishable population structure, all CipS cases and CipR controls were MLST and *spa* typed, revealing that CipS isolates were much more heterogeneous (Table 3.9, Figure 3.5). Comparisons between CCs of cases and controls showed that both Oxford and Brighton CipS isolates were significantly different from CipR controls (both $P<0.001$), although not different from each other ($P=0.13$). Only 3/21 Oxford CipS and 0/18 Brighton CipS were UK healthcare associated ST22 and ST36 as opposed to 20/21 of the CipR controls. Five Oxford CipS and eight Brighton CipS cases were ST1, and all other STs had frequencies of three or fewer, although seven Oxford and two Brighton CipS isolates were in CC5 (Table 3.9). *spa* types supported the results from MLST (Figure 3.5).

3.4.3.4 Microbiology with epidemiology

Epidemiological and microbiological information was combined to identify whether any CipS cases with STs other than healthcare associated ST22 and ST36 also had epidemiological characteristics in fitting with CA-MRSA. A group of particular interest was ST1 *spa* t127, which consisted of one third (13/39) of the CipS isolates. This was a distinct PVL negative strain that could not be grouped into a CC with any other MLST or *spa* type found in this study. However, individuals with ST1 *spa* t127 did not differ significantly from other CipS isolates in their epidemiological characteristics.

One Oxford and four Brighton strains carried PVL (Table 3.9), also used as a marker for CA-MRSA (section 1.4.3). Of the five PVL positive samples, three were isolated from individuals who had not been in hospital in previous year; and all five were isolated from either surface cultures (n=3) or pus samples (n=2). Two of the PVL positive strains, both from individuals with no previous hospital exposure, were ST80: a common European community associated strain (Elston and Barlow 2009). Additionally one PVL positive sample was ST8 *spa* t008, the same ST as community associated

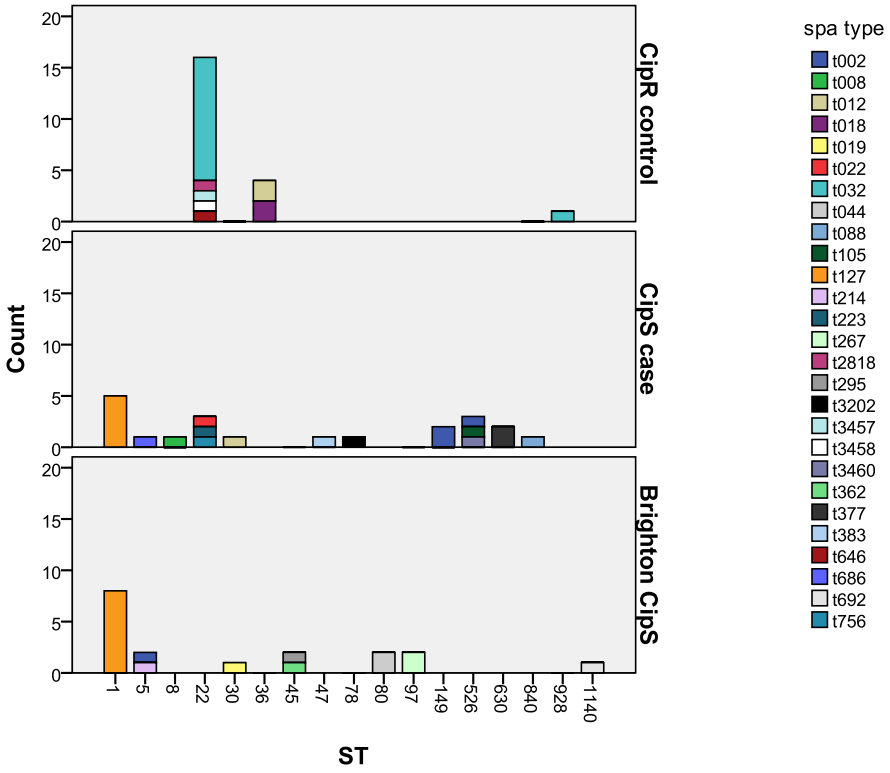
USA300 common in the USA, isolated from a surface culture from a 20 year old male, who had been in accident and emergency three days before isolation.

Table 3.9: MLST results of CipS cases and CipR controls

Strain marked with * is also a community case; strains marked with † carried the PVL gene.

CC	Sequence Type	CipS case (N=21)	Brighton CipS case (N=18)	CipR control (N=21)
Total		***21 (100%)***	***18 (100%)***	***21 (100%)***
CC5	***All STs***	***7 (33%)***	***2 (11%)***	***0 (0%)***
	5	1	2†	
	149 (1 SVL from ST5)	2		
	526 (1 SLV from ST5)	3		
	840 (1 SLV from ST5)	1*		
CC8	***All STs***	***3 (14%)***	***0 (0%)***	***0 (0%)***
	8	1†		
	630	2		
CC22	***All STs***	***3 (14%)***	***0 (0%)***	***17 (81%)***
	22	3		16
	928 (1 SLV from ST22)			1
CC30	***All STs***	***1 (5%)***	***1 (6%)***	***4 (19%)***
	30	1	1†	
	36 (1 SLV from ST30)			4
CC45	***All STs***	***1 (5%)***	***2 (11%)***	***0 (0%)***
	45		2	
	47 (1 SLV from ST45)	1		
CC88	***All STs***	***1 (5%)***	***1 (6%)***	***0 (0%)***
	78	1		
	1140 (1 SLV from ST78)		1	
CC1	1	5 (24%)	8 *(44%)*	
CC80	80		2†† *(11%)*	
CC97	97		2 *(11%)*	

Figure 3.5 STs separated by *spa* type for CipS cases and CipR controls



3.4.3.5 Identification of CA-MRSA

Investigation of individuals with CipS MRSA identified 11 cases with potential CA-MRSA out 39 initially investigated. These individuals had not been in hospital in the year prior to MRSA isolation, did not have healthcare associated STs and three had PVL. The finding that 28% of individuals with CipS MRSA were potentially CA suggests that ciprofloxacin sensitivity represents a highly useful definition when identifying potential CA-MRSA. However, detailed epidemiological information was not known about the CipS cases so it was not possible to determine whether they fit epidemiological definitions of CA-MRSA, such as the CDC definition. One of the 11 individuals identified was also a community case. Additionally, the use of a microbiological definition was able to identify a separate population of 25 individuals who did not have CA-MRSA but were clearly distinguishable from those with HA-MRSA. These individuals had recent hospital exposure but “wild” non-healthcare associated strains probably acquired in the community, and could again be termed community-associated MRSA.

3.5 Discussion

Across the three studies, 11 cases of CA-MRSA were identified with high confidence among 6801 individuals, in line with the initial part of the hypothesis that there is CA-MRSA in Oxfordshire, albeit at low levels. Investigation of the epidemiology of each set of cases compared to controls revealed that cases had distinct epidemiological characteristics. In all three studies cases were significantly younger than controls, and the median age of the 11 individuals identified with CA-MRSA was 37 years compared to 75 for all other individuals included in the three studies (P=0.0005). CA-MRSA also had a distinguishable population structure as the 11 CA-MRSA cases identified all had STs other than epidemic hospital strains ST22 and ST36. Results from the three case-control studies are therefore consistent with the hypothesis that: there is community-acquired MRSA in Oxfordshire which has a distinguishable epidemiology and population structure.

This study shows that true CA-MRSA constitutes less than 1% of MRSA in Oxfordshire. Samples were taken from the entire Oxfordshire region, so it is likely that results can be generalised to other UK counties, confirming that currently CA-MRSA is not causing the health threat seen in the USA. The low level of CA-MRSA found in this study is similar to that reported later from a review of studies throughout the UK in 2009 (Elston and Barlow 2009), suggesting that, since this study levels of CA-MRSA have not increased, and further supporting generalisation of the study results throughout the UK.

This study found that younger age and population structure, as defined by MLST and *spa* typing, distinguished CA-MRSA. Although CA-MRSA was previously known to be associated with younger age (Diederen and Kluytmans 2006), these findings are from work performed outside the UK and the association has never been confirmed in a UK cohort not specifically linked to an outbreak. Similarly, investigation of outbreaks is likely to identify an outbreak specific population structure, therefore confirmation that CA-MRSA has a different population structure in a general (non-outbreak) sample from the UK population is novel to this study.

Each of the 11 CA-MRSA isolates identified were either surface cultures (n=7) or pus samples (n=4), despite being taken from collections that included screens and bacteraemia samples. This confirms

that CA-MRSA in the UK has a similar phenotype to that found outside the UK (Diederen and Kluytmans 2006), and is therefore likely to cause disease in the form of skin and soft tissue infection. In particular, only 8% of CipS cases were screening samples which suggests that the majority caused disease, highlighting the importance of identifying and preventing MRSA in the community.

3.5.1 Compromises in study design

Each of the three studies represented compromises in the cases selected and information available. In particular, a limitation of any case-control study is that it relies on written historical data. The information in this study was gathered from admission data, hospital notes, GP records and GP questionnaires. It is therefore is dependent upon what was recorded (recall bias) as well as the accuracy of the information given by the patient (reporting bias).

Further compromises were specific to each of the studies. Cases from study one: individuals with MRSA admission bacteraemia, were defined only by current hospital admission to prevent exclusion of individuals with community acquisition despite previous hospital contact. However, this was not a sufficiently rigorous definition to identify CA-MRSA, since 97% of cases had been in hospital in the previous year. Additionally, bacteraemia is not the most common form of *S. aureus* disease, since CA-MRSA largely presents as skin and soft tissue infection (Elston and Barlow 2009), so individuals with CA-MRSA on admission and a presentation other than bacteraemia may have been missed.

Case-control study three identified individuals using a microbiological definition. Due to the difficulty of collecting accurate epidemiological information, detailed epidemiology was not available about CipS cases and CipR controls, therefore the 10 individuals with CA-MRSA identified only as CipS cases may have had other unknown healthcare risk factors. However, epidemiological definitions do not always accurately identify individuals with CA-MRSA, as people with healthcare exposure may acquire MRSA elsewhere. It was for this reason that case-control study one identified cases as any individuals with MRSA bacteraemia on admission. In support of this, one individual identified with CipS MRSA, had been to accident and emergency three days before MRSA isolation, but had no other hospital exposure in the previous year. This individual had ST8 *spa* t008, synonymous with

community associated strain USA300, which was likely acquired before the visit to accident and emergency, however, under epidemiological definitions would not be described as CA-MRSA.

3.5.2 Impact of results

The primary impact of this study is that CA-MRSA represents less than 1% of UK MRSA. Therefore, in the UK, individuals without healthcare risk factors are unlikely to carry MRSA. This has implications for treatment of patients on admission to hospital and confirms that future outbreaks arising from a reservoir of potentially more virulent MRSA in the community spreading into hospital are unlikely.

Whilst proving that CA-MRSA levels are low in the UK, artefacts of the definitions used in this study also identified two groups of community-associated MRSA which although not traditionally community-acquired, have important implications. Firstly, “feral” MRSA was found in hosts with no healthcare risk factors but with molecular characteristics identical to hospital-associated strains. “Feral” MRSA may represent prolonged carriage of MRSA acquired in hospital over a year before; or MRSA acquired during other non-hospital healthcare, such as identified from GP questionnaires in case-control study two. This has implications for infection control policies, since individuals without hospital risk factors may still carry community-associated MRSA which could be transmitted if admitted to hospital.

Secondly “wild” strains were identified in individuals with healthcare risk factors, but not typical nosocomial MRSA types. “Wild” strains are potentially important as they may represent a source for clonal expansions producing future epidemics, such as USA300 in the USA. An example of this could be ST1 *spa* t127, which was the fourth most common ST found in this study and has been shown to be increasing in prevalence in a study of drug users in Brighton (Atkinson, Paul et al. 2009). Such distinctions around healthcare-associated MRSA have never been made before, and are of use for future investigation of MRSA in the community.

Investigation of the hypothesis using three definitions meant that study results can also be used to evaluate of definitions of CA-MRSA, an important factor for UK infection control policy. Case-control study one used an origin based definition and found no CA-MRSA, suggesting that such definitions are not sufficient to identify CA-MRSA, despite the advantages in their simplicity. The rigorous

epidemiological definition used in case-control study two found one individual with CA-MRSA, however required detailed investigation of 105 individuals, each of whom had their GPs contacted. Finally case-control study three identified 11 individuals from 39 investigated, suggesting that ciprofloxacin sensitivity is a good marker for CA-MRSA. These individuals would require further epidemiological information to confirm that they had no other healthcare exposure, however results of studies performed after this study confirm the value of using ciprofloxacin sensitivity for identifying CA-MRSA (Otter and French 2008; Popovich, Weinstein et al. 2008).

After testing each of the three main definitions of CA-MRSA described at the beginning of this chapter it is possible to suggest the most effective way to identify CA-MRSA. The best definition of CA-MRSA would consider factors from case-control studies two and three, i.e. microbiological information alongside careful use of healthcare risk factors. Such a complex definition is not practical to identify CA-MRSA, however, a staged version would be possible. Here, isolates could firstly be identified using the simple but discriminatory microbiological criterion of ciprofloxacin sensitivity, following which, more detailed epidemiology could be considered on a smaller group of samples. This would include rigorous investigation of the intensity and frequency of healthcare risk factors, to fully confirm whether an isolate is community-acquired.

3.5.3 Future work

To confirm the low level of true CA-MRSA found in this study it would be necessary to investigate samples isolated in the community which were not sent into hospital. This would require investigation of members of the general public who may not have MRSA disease, but could be healthy MRSA carriers. Such a study could also include investigation of factors not included in healthcare records, such as social behaviour and family contacts, which may also be risk factors for CA-MRSA. As well as investigating the hypothesis of this chapter in more detail, a study investigating MRSA in the community would be able to answer hypotheses such as: CA-MRSA is present in the community in isolates not sent into hospital for testing; and CA-MRSA in Oxfordshire has distinguishable non-healthcare risk factors.

This study identified ciprofloxacin sensitivity as a discriminatory marker for CA-MRSA, however, detailed epidemiological information was not available to confirm whether individuals identified had true CA-MRSA. To fully support this finding, it would be necessary to combine the definitions used in case-control study two and three as recommended above. This would determine whether CipS MRSA truly represents CA-MRSA and enable hypotheses such as: ciprofloxacin sensitivity is a marker for CA-MRSA, to be investigated.

Finally, MLST and *spa* typing are relatively crude measures for investigation of the population structure of CA-MRSA. Whole-genome sequencing could offer a much more detailed investigation of CA-MRSA population structure, including investigation of genetic differences between “feral” and hospital acquired strains, identical using common molecular typing techniques. This would enable investigation of hypotheses such as: nosocomial MRSA is genetically distinct from “feral” MRSA, and could enable characterisation of strains acquired in the community that are currently believed to be the same as hospital types.

4 Population structure and carriage length of *Staphylococcus aureus* in Oxfordshire

4.1 Introduction

A prospective study was designed to investigate *S. aureus* carriage in individuals from the community and the host characteristics and pathogen genetics associated with this. Individuals were followed over a two year period to investigate *S. aureus* dynamics over time and its correlation with healthcare behaviour. This study allowed investigation of *S. aureus*, including MRSA, in the community in more depth than any previous study, advancing on previous chapters in this thesis, which used only samples sent for routine hospital microbiology testing and details obtained from hospital and GP records.

4.2 Hypothesis

The study was designed to investigate the following hypothesis:

> *S. aureus* carriage in the community is correlated with environmental host risk factors and bacterial genetic factors

In addition, the study was designed to further investigate part of the hypothesis from Chapter Three: there is CA-MRSA in Oxfordshire.

4.2.1 Importance of the hypothesis

S. aureus carriage is a well known risk factor for disease, so a detailed investigation of carriage in the community will benefit our understanding of *S. aureus* transmission and its progression to disease. Additionally, *S. aureus* in the community is a reservoir for transmission into hospital, hence it is vital to understand in order to reduce the burden of nosocomial *S. aureus* infections. Despite its importance, little is known about healthy carriage of *S. aureus* in the UK, since most studies investigating *S. aureus* focus on specific populations with increased risk of disease.

The study investigated *S. aureus* carriage using two methods, firstly cross-sectional carriage, was investigated by comparing individuals positive and negative for *S. aureus* from a single nasal swab.

Secondly, longitudinal carriage was investigated by regular nasal swabbing over a period of two years. Host environmental risk factors were correlated with each measure of carriage, thus enabling identification of individuals more likely to carry *S. aureus* and those likely to carry for longer who are at greater risk of progression to disease. Identification of host factors associated with increased risk of carriage and longer term carriage is important, since it would allow healthcare to be altered accordingly for those at risk. Investigation of longitudinal carriage is also important to identify individuals who may acquire *S. aureus* in hospital, but then carry for a long period, possibly acting as vectors transmitting *S. aureus* acquired in hospital to others in the community.

In addition to the correlation of host factors with *S. aureus* carriage, each swab positive for *S. aureus* was *spa* typed, including multiple *spa* types for samples that showed signs of multiple strains (section 4.3.3.4). This enabled correlation of *S. aureus* genetics with carriage behaviour over time, allowing potential associations with strains that were more likely to be carried for longer durations. Identification of *S. aureus* strains with certain characteristics could allow strain-specific treatment, given to patients on the basis of how long their strain is likely to colonise them for. Additionally, once strains with certain characteristics, such as prolonged carriage, are identified they can be further investigated, perhaps by whole-genome sequencing and functional studies, in order to determine the cause of these characteristics.

4.2.2 Novelty of the hypothesis

Although previous studies in the UK have investigated risk factors associated with cross-sectional *S. aureus* carriage, they are either considerably smaller than this study, or investigate specific sub-populations (Abudu, Blair et al. 2001; Peacock, Justice et al. 2003; Nulens, Gould et al. 2005). This study will generate reliable estimates of general UK *S. aureus* carriage rates, which are otherwise not available, and correlate them with detailed host factors.

This study includes the collection of detailed data on risk factors for *S. aureus* carriage from multiple sources, including participant interview, GP and hospital records. Although risk factors for *S. aureus* carriage have been investigated before, studies are often focussed on specific populations (Bischoff, Wallis et al. 2004; Herwaldt, Cullen et al. 2004; Halablab, Hijazi et al. 2010; Qu, Cui et al. 2010); are

not specifically designed to investigate *S. aureus* (Graham, Lin et al. 2006; Kuehnert, Kruszon-Moran et al. 2006); or do not have access to medical records (Miller, Cook et al. 2009; Munckhof, Nimmo et al. 2009). Therefore, this study will be the first to relate social, demographic and healthcare related risk factors to *S. aureus* carriage.

In addition to a more comprehensive investigation of cross-sectional *S. aureus* carriage in the UK, this study also provides in depth investigation of carriage over time. Previous studies on longitudinal *S. aureus* carriage largely focus on MRSA and individuals previously known to be carriers, with durations of follow-up often shorter than one year (Sanford, Widmer et al. 1994; MacKinnon and Allen 2000; Vriens, Blok et al. 2005). Even fewer previous studies have correlated carriage with risk factors, and in those that do follow-up durations are short and limited risk factor data is collected (Harbarth, Liassine et al. 2000; Scanvic, Denic et al. 2001; Marschall and Muhlemann 2006; Lucet, Paoletti et al. 2009). Finally, no previous studies consider whether the strain of *S. aureus* carried has changed over time, including whether individuals' positive for *S. aureus* carried multiple strains, or whether individuals who were initially negative acquired new strains. This study was able to investigate the acquisition rate in individuals initially negative for *S. aureus* as well as the proportion of people who change strain during carriage, both of which were correlated to host risk factors. Such detailed investigation of *S. aureus* longitudinal carriage has never been carried out before.

4.3 Methods

The *S. aureus* carriage study was designed to investigate healthy individuals in the community for both cross-sectional and longitudinal *S. aureus* carriage. Participants followed longitudinally were a subset of those initially investigated cross-sectionally. To investigate host factors, participants were interviewed and further information was taken from GP and hospital records. To investigate bacterial genetic factors, each swab positive for *S. aureus* was *spa* typed, including investigation of multiple colonisation in the nose at one time.

4.3.1 Participant recruitment

Participants were recruited from five practices: Summertown Health Centre; Moorland House Surgery, Wheatley; Horsefair Surgery, Banbury; The Health Centre, Bicester; and Bampton Medical

Practice. These practices were chosen as they are part of the Thames Valley Primary Care Research Partnership, which is part of the National Institute of Health Research. To increase numbers of young participants, additional students registering at Summertown Health centre were recruited during the University Freshers Week. Heather Godwin recruited the majority of participants. Information sheets and consent forms given to participants are shown in Appendix 4 and 5.

The decision to recruit participants from GP surgeries was made, since they offered a very useful and convenient combination of circumstances. Firstly, the research nurse was able to opportunistically approach participants who had time to read the information sheet about the study whilst waiting for their appointment. The nurse could be in a private room which participants could approach immediately after their appointment, rather than being asked to attend a separate meeting to be recruited, which improved recruitment rates, especially in younger people who have been shown to be less likely to respond when asked to attend a GP surgery for a similar research study (Abudu, Blair et al. 2001). Secondly, recruitment from GP practices made it easier to collect medical information from participants GP records after recruitment.

It is possible that individuals visiting their GP may over select people in poor health, and therefore not properly represent the general population without underlying disease. However, previous research on *S. aureus* carriage in individuals recruited from GP practice compared to others selected from the electoral role, found no difference in risk factors for *S. aureus* carriage between the two groups (Munckhof, Nimmo et al. 2009). It must be noted that Munckhof *et al* recruited individuals only if they were presenting with non-infectious conditions, which was not possible in this study as there was not ethical approval to ask potential participants the reason for visiting their GP. However, to gain a more accurate representation of the general population, it was aimed to stratify the sample to match the age and sex distribution of the UK general population above 15 years old.

4.3.1.1 Exclusion criteria

Due to the different nature of *S. aureus* carriage in children, individuals under 16 years old were not recruited into the study. Children do not have fully developed immune systems and therefore have a different host organism relationship from adults. With a sample size of 1000 the number of children

included in the study would be too small to provide sufficient power to make conclusions about *S. aureus* carriage and disease progression in children.

Additionally, as the study aimed to investigate progression of *S. aureus* carriage over time, people were not recruited into the study if they were unable to comply with follow-up. This included individuals unable to consent or individuals anticipating missing more than two follow-up swabs. In contrast to children, individuals unable to consent, or otherwise participate in follow-up, are unlikely to have a different host organism relationship from other adults, therefore their exclusion would not affect the study results.

4.3.2 Numbers recruited

Power calculations were performed to determine the number of participants needed to detect differences between carriers and non-carriers and intermittent and persistent carriers. When calculating power it was assumed that 35% of individuals in the study would carry *S. aureus* and that one third of these would be persistent carriers. Calculations were made based on frequency of risk factors established to be significant in previous studies for *S. aureus* carriage versus non-carriage (Table 4.1) and for persistent and intermittent carriage (Table 4.2).

Table 4.1: Power to detect significant risk factors found in previous studies with a sample size of 1000 and an estimated carriage rate of 35%

Power above 0.8 is considered sufficient (marked in bold)

Risk factor	Percentage of *S. aureus* carriers with risk factor	Percentage of *S. aureus* non-carriers with risk factor	Power	Reference
Male sex	41%	31%	**0.85**	(Miller, Cook et al. 2009)
Male sex	40%	30%	**0.90**	(Bischoff, Wallis et al. 2004)
Antibiotic use in previous month	4%	10%	**0.94**	
Male sex	41%	30%	**0.94**	(Munckhof, Nimmo et al. 2009)
Age >60	31%	41%	**0.87**	
Caucasian ethnicity	89%	79%	**0.98**	

Table 4.2: Power to detect significant risk factors found in previous studies with a sample of 360 carriers and an estimated rate of 33% persistent carriage

Power above 0.8 is considered sufficient (marked in bold)

Risk factors	Percentage of persistent carriers with risk factor	Percentage of intermittent carriers with risk factor	Power	Reference
Antibiotics	41%	23%	**0.95**	(Marschall and Muhlemann 2006)
Immunosuppressive therapy	57%	28%	**0.99**	
Haemodialysis	80%	30%	**1.00**	
Skin break	55%	18%	**1.00**	(Scanvic, Denic et al. 2001)
Self sufficiency in daily activities	37%	30%	0.27	(Lucet, Paoletti et al. 2009)
Female sex	55%	25%	**1.00**	(Harbarth, Liassine et al. 2000)
Recent fluoroquinolones	19%	28%	**1.00**	

According to the calculations in tables 4.1 and 4.2 a sample size of 1000 individuals from the general population is sufficient to detect differences between carriers. With an estimated rate of carriage of 35%, this sample size would detect all the previous differences observed between carriers and non-carriers and 6/7 of the previously observed differences between intermittent and persistent carriers if one third of carriers carried persistently.

4.3.3 Methods for swabbing and culture of *S. aureus*

4.3.3.1 Nasal swabs

Swabs were taken from the anterior nares, since that is where *S. aureus* is most consistently isolated (Williams 1963; Sanford, Widmer et al. 1994; Sivaraman, Venkataraman et al. 2009) and because nasal swabs are easy to administer, allowing self-swabbing during study follow-up. Charcoal swabs were used as previous studies have shown this is the most effective way of taking and storing nasal samples (Eriksen, Espersen et al. 1994; Eriksen, Espersen et al. 1995). Swabbing only the nose might have missed some individuals who carried *S. aureus* elsewhere in the body, such as the throat or groin, and therefore slightly underestimate carriage rates. Further arms of the project this study forms part of, will collect clinical isolates from other parts of the body as well as nasal swabs, so comparison of *S. aureus* between these sites will be possible.

4.3.3.2 Pilot studies

To decide on the best mediums for swabbing, transporting and culturing nasal swabs, two pilot studies were performed. The first aimed to determine whether a wet charcoal swab was as likely to grow *S. aureus* as a dry charcoal swab; whether enrichment increased the chance of a swab growing *S. aureus;* and whether CBA, Manitol Salt Agar (MSA) (Bio-Rad, Limerick, Ireland) or chromogenic agar was more likely to grow *S. aureus*. The pilot study included only culture based methods, since these represent standard *S. aureus* identification (section 1.1) and allow isolates to be easily stored for future investigation. It is possible that use of PCR based techniques would have indentified individuals with very low levels of *S. aureus* not detected in routine culture.

Eleven volunteers were asked to swab each nostril according to the instruction leaflet sent to participants. Participants swabbed each nostril with a separate swab, one wet charcoal swab, placed in saline solution before the nostril, and one dry charcoal swab. Swabs were immediately cultured onto CBA, then placed in 5% NaCl broth and incubated at 37 ^{o}C overnight before being cultured onto three different agars, CBA, MSA, and chromogenic agar on which *S. aureus* grows as pink colonies.

Four of 11 individuals included in the study had a positive result on at least 1/4 agars (Table 4.3). No samples were meticillin resistant. All four individuals were positive for *S. aureus* on MSA and

SaSelect agar. For individual D only the wet swab was *S. aureus* positive. However, this may have been because individual D had a furuncle under the nostril the wet swab was taken from and on further investigation after the furuncle had gone, *S. aureus* could not be grown from either nostril with a wet or dry swab. It was decided to use overnight enrichment in 5% NaCl broth before culturing the strains onto chromogenic agar, since enrichment increased the likelihood of growing *S. aureus* and chromogenic agar was the easiest method to detect *S. aureus* and did not have a difference in culture sensitivity. These findings support other studies (Eriksen, Espersen et al. 1994).

Table 4.3: Results from a pilot study comparing growth of *S. aureus* from wet and dry nasal swabs cultured before and after enrichment on CBA, MSA and SaSelect agar

Individual	Wet or Dry	*S. aureus* detected on the following:			
		CBA from initial swab	CBA after enrichment	MSA after enrichment	SaSelect agar after enrichment
B	Dry	Negative	Positive	Positive	Positive
	Wet	Negative	Positive	Positive	Positive
D	Dry	Negative	Negative	Negative	Negative
	Wet	Positive	Negative	Positive	Positive
G	Dry	Positive	Positive	Positive	Positive
	Wet	Positive	Positive	Positive	Positive
J	Dry	Positive	Positive	Positive	Positive
	Wet	Positive	Positive	Positive	Positive

Follow-up required swabs to be returned to the hospital by post, therefore a second pilot study was undertaken to determine whether *S. aureus* can still be detected if culture is delayed. Here, three of the volunteers identified with *S. aureus* in their nose in the first pilot were asked to swab each nostril twice using dry charcoal swabs. Swabs were randomly assigned to be cultured after being left at room temperature for one, three, seven and eight days or being sent through Royal Mail second class post. After the specified delay, swabs were cultured according to the protocol from pilot study one. All swabs grew *S. aureus*, so the method was considered to be appropriate for identifying *S. aureus* after swabs had been sent by post. Complete details of the finalised method for culture of swabs returned as part of the study are provided in the Methods.

4.3.3.3 Routine culture

Due to the clinical significance of MRSA, swabs that tested positive for MRSA during the study were forwarded to the National External Quality Assessment Service (NEQAS) accredited NHS laboratory for confirmation. Results that confirmed a sample was positive for MRSA were sent to GPs and the relevant infection control team was notified.

4.3.3.4 *spa* typing

All samples positive for *S. aureus* were *spa* typed according to the Methods. *spa* typing alone was chosen, rather than MLST as used in Chapter Three of this thesis, since it is more discriminatory than MLST and requires sequencing of only one gene. A large number of samples were typed, so sequencing only one gene allowed considerable reductions in time and cost.

DNA for *spa* typing was extracted from a mixed sweep of colonies, allowing for detection of multiple *S. aureus* strains from one nasal swab (section 2.3.1.2). Nasal swabs that appeared to have mixed *S. aureus* growth were identified from chromatograms with either clear double peaks or unreadable sequence. (Those with unreadable sequence were initially sequenced again after a second DNA extraction to ensure the unreadable sequence was not a product of bad quality DNA used for sequencing.) All samples with chromatograms indicative of mixed colonisation were re-cultured and 12 individual bacterial colonies selected per sample and individually *spa* typed. Validation that chromatograms with clear sequence consisted of a single *spa* type was performed by *spa* typing 24 single colonies from 32 samples that yielded clear chromatograms, which all produced identical *spa* types. Methods and laboratory procedures for detection of multiple strains from one swab were designed and performed by Dr Antonina Votintseva.

4.3.4 Collection of risk factor information

The study provided a unique opportunity to investigate a large sample of participants. It was therefore important to design a comprehensive investigation of risk factors for *S. aureus* carriage; carriage duration; disease progression and presentation; and carriage acquisition. The participant questionnaire was used to collect basic demographic information, as well as more detailed personal information about factors possibly related to *S. aureus* carriage, such as employment, household

contacts, leisure activities, travel and care for individuals who may be at increased risk of *S. aureus* carriage. In addition, the participant questionnaire provided an opportunity to collect healthcare information that might not be recorded reliably in GP or hospital records, such as receiving personal care or having a urinary catheter.

Further healthcare risk factors were collected from GP records, such as appointments with GPs or nurses, any hospital appointments outside ORH, chronic healthcare conditions, other co-morbidities and any antibiotics and steroid. The term "long term illness" was used to encompass any chronic illnesses which were each anticipated to give a small increased risk of *S. aureus* but to be present in a small minority of participants, so would not have reached statistical significance individually. This approach is similar to Chapter Three, where various illnesses were grouped into a co-morbidity score. Details of individual long term illnesses were documented, allowing further investigation should long term illness be a significant predictor of any dependent variable. Finally, hospital appointments inside ORH were taken from ORH electronic records. After completion of the study at two years, the study nurse will return to GP and ORH electronic records to collect the same information about participants for the two years during which they were participating in the study. A full list of information collected for each participant is given in table 4.4 and questionnaires completed from participant interview and GP records are included as Appendix 6.

Table 4.4: All risk factors collected during the study

Risk factor	Information obtained from	Reference
Basic demographics:		
Age	Participant interview	Admission bacteraemia study, Community-MRSA study
Sex	Participant interview	Admission bacteraemia study, Community-MRSA study
Ethnic background	Participant interview	(Bhattacharya, Carleton et al. 2007)
Place of residence	Participant interview	Admission bacteraemia study, Community-MRSA study
Employment:		
Current employment	Participant interview	This study
Healthcare related employment	Participant interview	Community-MRSA study
Household:		
Members of usual household	Participant interview	(Davis, Perri et al. 2007)

Risk factor	Information obtained from	Reference
Members with healthcare contact	Participant interview	(Davis, Perri et al. 2007)
Leisure:		
Sport	Participant interview	(Davis, Perri et al. 2007)
Travel	Participant interview	(Davis, Perri et al. 2007)
Healthcare contact:		
In-patient: most recent, and all in previous year	Hospital and GP records	Admission bacteraemia study
Out-patient: most recent, and all in previous year	Hospital and GP records	Admission bacteraemia study
GP visits: most recent, and all in previous year	GP records	Community-MRSA study
Practice nurse visits: most recent, and all in previous year	GP records	Community-MRSA study
District nurse visits: most recent, and all in previous year	GP records	Community-MRSA study
Any other help at home: e.g. with shopping, meals on wheels, cleaning	Participant interview	(Davis, Perri et al. 2007)
Co-morbidities:		
Long term illness	GP records	Admission bacteraemia study
Healthcare history:		
Renal dialysis	GP records	Admission bacteraemia study
Chemotherapy	GP records	(Graffunder and Venezia 2002)
Surgery	GP records	(Rezende, Blumberg et al. 2002)
Oral steroids	GP records	(Lodise, McKinnon et al. 2003)
Vascular access	Participant interview	Admission bacteraemia study
Insertion of urinary or suprapubic catheter	Participant interview	Admission bacteraemia study
Treatment for skin problems possibly linked to *S. aureus* in previous month*	GP records	Community-MRSA study
Percutaneous vascular access in previous month	GP records	Admission bacteraemia study
Antibiotics: most recent, and all in previous year	GP records	Admission bacteraemia study
Previous MRSA	GP records	Admission bacteraemia study
Previous MSSA	GP records	This study
Other activities:		
Contact with at risk: mental ill health or disability or elderly	Participant interview	(Davis, Perri et al. 2007)
Receiving domestic help or social service support	Participant interview	(Davis, Perri et al. 2007)
Personal care at home	Participant interview	(Davis, Perri et al. 2007)

* Skin problems include trauma, cut, splinter, cannula, arterial or venous catheter, surgery, eczema, psoriasis, ulcer, rash (e.g. chicken pox, herpes, zoster), insect bite, bite, tattoo or piercing.

4.3.4.1 Management of data

Data from participant interviews and GP records was completed as a paper questionnaire by the study nurses; data from hospital records was gathered electronically. Information was entered into a Microsoft Access database constructed for the study, located on the ORH trust server. The database was designed so that name, address, postcode, hospital number, NHS number and clinician could be removed for data analysis outside the ORH server. Paper forms were subsequently stored in a locked filing cabinet. Each participant was given a unique identifying number to link them to any samples or further analysis separated from other identifiers.

4.3.4.2 Unacceptable values

Information from the database was imported into STATA/IC Version 11.1 (StataCorp LP, USA) and checked for unacceptable values, particularly impossible dates (for example date of most recent GP appointment recorded after recruitment date). When incorrect data was identified the information was checked on the paper copy of the questionnaire. In the unusual circumstance that the data was incorrect on the paper questionnaire the result was discussed with the study nurse, and where possible, the nurse returned to the original records to check.

4.3.4.3 Missing values

It was inevitable that some information would not be obtainable for all participants. This was particularly the case with information from GP records which are often incomplete, especially if the individual has recently moved to their GP practice. For all questions requiring a Yes/No answer there was also an "unknown" option and for dates, those which were anticipated to be uncertain (such as dates of travel) were entered as days, months and years separately so unknown units of time could be left blank, avoiding approximation of dates. For other dates when values were uncertain, days were rounded to the 15th of the month or 1st July where month was unknown. Individuals with many missing values were removed from analysis.

4.3.4.4 Data consistency

Four people were involved in completing questionnaires. To ensure consistency in data entry between researchers, weekly meetings were held to compare data entered into the questionnaire and

clarify terms used. For example “contact sport”, “healthcare related employment” and “long term illness” were discussed to ensure that different investigators were entering comparable data. Additionally, small modifications were made to the questionnaire as the study progressed to ensure all relevant information was collected in the simplest way possible. For example, information collected from GP records was modified by addition of more detailed information about skin breaks in the month prior to baseline; the source of information about previous catheters and vascular access was changed to the participant questionnaire as it was not reliably recorded in GP records; and the form was rearranged so that it was quicker and easier to complete.

4.3.5 Follow-up

Initially, all patients with a positive nasal swab at baseline were included in follow-up. Later in the study, an amendment to the protocol allowed the inclusion of participants with an initial negative nasal swab to investigate *S. aureus* acquisition. These individuals included all the participants from Bampton Medical Practice and students recruited in University Freshers Week.

All individuals participating in follow-up were sent a nasal swabbing kit through the post one month after recruitment, then every two months for two years. Follow-up by post represented a practical method to obtain regular swabs from all participants enrolled in follow-up, however, required an element of trust that participants swab their own noses properly. The swabbing kit included an information leaflet, details of the participant’s previous result if requested, and a short form to complete giving details of any antibiotics the participant may have been prescribed by anyone other than their GP (information about antibiotics prescribed by a GP will be collected when revisiting GP notes after two years) (Appendix 7). At the initial GP visit, the nurse explained how to take and send the swab and showed the participant the contents of a swabbing pack. Instructions were also repeated in the leaflet provided with each swabbing kit (Appendix 8). With every swabbing kit sent, participants were asked to swab themselves and send back a charcoal swab and completed form using an enclosed stamped envelope and packaging.

4.3.5.1 Adherence

Adherence to swabbing was monitored and patients reminded by phone, email and text message where appropriate (up to two reminders, three and six weeks after due date). Care was taken to ensure no distress was caused if the participant was deceased by checking with their GP before contact was made. If a participant did not provide a swab after this time, they were contacted again for the next swab. Participants were removed and their data censored after the point of removal if they failed to return three swabs, if they requested to be removed from the study or if they moved out of the country.

4.3.5.2 Clinical *S. aureus* isolates

During follow-up all clinical *S. aureus* isolates from study participants sent to the ORH microbiology laboratory were identified. To do this a daily electronic report of clinical *S. aureus* samples was produced by the ORH microbiology laboratory. These isolates were stored for use in future analysis comparing carriage and clinical *S. aureus* strains.

4.3.6 Statistical methods

Statistical analyses were performed using Stata/IC Version 11.1. To describe continuous data, medians and inter-quartile ranges were used as they are not skewed by extreme variables. Univariate analysis was performed using Mann-Whitney rank sum tests for continuous variables and Fisher's Exact Test to compare categorical variables, except for analysis of carriage groups (section 4.4.4.2) where multinomial logistic regression was used, since this is a powerful way to test a dependent variable with four categories. Time-to-event outcomes were analysed using Kaplan-Meier plots and log rank tests. *P* values <0.05 were considered statistically significant.

Multivariate analyses were performed to consider the effect of multiple variables at once, using multinomial fractional polynomial (MFP) regression with four degrees of freedom, and the Akaike Information Criterion (AIC) for model comparison. The fractional polynomial method allowed for a non-linear relationship between dependent and independent variables, with four degrees of freedom equivalent to a quadratic relationship used where necessary (otherwise one degree of freedom was used). MFP regression removed variables in a backward stepwise manner in order to obtain the most

powerful model. The type of regression model used was dependent on the nature of the dependent variable. Logistic regression was used for dichotomous independent variables; multiple logistic regression was used for categorical independent variables; and Cox regression was used when the dependent variable was a time-to-event outcome. Where there were co-dependencies between variables, only one measure was entered into the multivariate model.

MFP regression analysis based on backwards selection using the AIC was the multivariable method chosen to analyse the data, since it allowed investigation of any potentially non-linear relationships between multiple dependent continuous and categorical variables (exposures) and the independent variable (outcome) using a statistical method which controls for multiple testing using an approximation to a closed testing procedure (Sauerbrei 2008). Use of multivariate as opposed to univariate regression analysis reduced the false attribution of associations by accounting for underlying relationships between the independent variables which may affect their relationship with the dependent variable, i.e. adjusted for confounding bias. However, regression analysis can only ascertain relationships and does not necessarily infer the underlying causal mechanism. Therefore an association between two variables does not necessarily imply the direction of causality from the independent to the dependent variable. Additionally, when there are two reasonably highly correlated independent variables it is possible that the variable with the true causal effect could be incorrectly removed from the model as a chance consequence of variation due to sampling, leaving a co-correlated but actually non-related explanatory variable in the final model. As well as this, factors not included in the model could actually be unmeasured confounders if correlated with other dependent variables and the outcome; that is, any associations that are identified in the final model are valid only under an assumption of no unmeasured confounders. Therefore, when interpreting the results of the regression analysis, care must be taken to emphasise that although a relationship has been identified, it is not necessarily causal; may be specific to the sample dataset; and relies on all relevant factors having been measured and appropriately adjusted for in the model.

Due to the requirement of collecting detailed information from GP and hospital records, it was inevitable that not all information would be available for every participant. Missing data was managed

in a consistent manner to ensure it did not affect results. For categorical variables, missing data was given a separate category and included in the analysis. For time-to-event variables with <25% missing, missing values were imputed as the 99th percentile for that value and included in the analysis, since missing data could reasonably be assumed to indicate that the event had not occurred recently (as otherwise it would have been recorded with high probability). Where >25% of data was missing, it was not sensible to estimate the missing data, and therefore time-to-event variables were converted into categories with an additional category for missing data.

4.3.7 Ethical approval

The study was approved by Oxfordshire Research Ethics Committee B on 8 September 2008. The reference number is 08/H0605/102.

4.4 Results

4.4.1 Participant recruitment, data collection and compliance

4.4.1.1 Recruitment

The study aimed to recruit 1000 participants, 200 from each of five GP practices, with an age and sex distribution matching that of the general population aged above 15 years (Office for National Statistics 2007) (Table 4.5). Between 17 December 2008 and 2 December 2009, 1123 participants were recruited, including 121 university students recruited to re-balance the age distribution. Because of age and sex specific patterns of GP attendance and lower recruitment success in younger people, it was not possible to exactly match the population age and sex distribution; however each age/sex group (defined by the Office for National Statistics (Office for National Statistics 2007)) included more than 50 individuals.

Table 4.5: Recruitment by age and sex compared to general population

Age/Sex group	Number in study	Number per 1000 general population age sex distribution	Percentage of number required for parity
16-29 Male	87	113	77%
16-29 Female	126	110	115%
30-44 Male	59	137	43%
30-44 Female	118	139	85%
45-59 Male	97	118	82%
45-59 Female	166	121	137%
60-64 Male	61	31	197%
60-64 Female	75	33	227%
65-74 Male	98	49	200%
65-74 Female	98	55	178%
75+ Male	77	36	214%
75+ Female	61	58	105%
TOTAL	1123	1000	

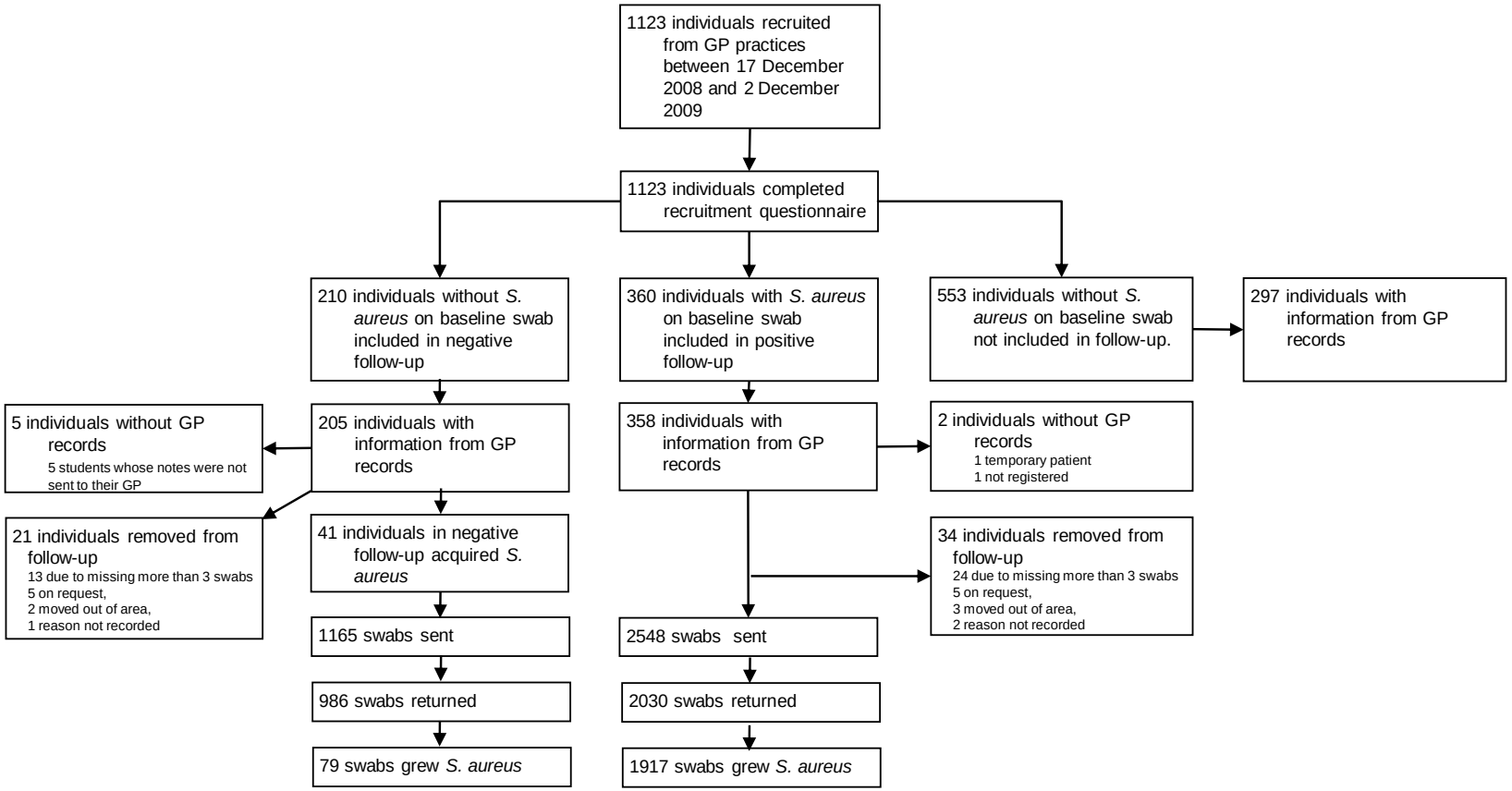
4.4.1.2 Data collection

The study had three arms, differentiated by the result of the first swab for each participant. Firstly, all participants were investigated for cross-sectional carriage, after which participants who had an initial *S. aureus* positive swab (nasal colonisation at baseline) (360/1123 (32%)) were enrolled in a group referred to from herein as “positive follow-up”. In order to investigate *S. aureus* acquisition, 210 initial swab-negative individuals were enrolled in “negative follow-up” (all participants from Bampton Medical Practice and students from Summertown Health Centre). Five-hundred and fifty-three participants, negative at baseline were not included in any follow-up (Figure 4.1).

In order to collect information on host environmental factors two questionnaires were completed: the first by the study nurse upon recruitment, and the second using information from GP records. Additionally, ORH hospital exposure data was collected from ORH electronic records for all participants using a database query. The recruitment questionnaire was completed for all 1123 individuals. GP record information was collected for 358/360 individuals with *S. aureus* (the two missing were a temporary patient and one that had not registered at the surgery) and 205/210 individuals in negative follow-up (five missing were all students whose notes had not been sent to the GP practice). To improve power to detect predictors of initial *S. aureus* status, GP record information

was also collected from 297 controls negative for *S. aureus* who were not included in negative follow-up (controls were distributed evenly across the five GP practices and randomly selected within each).

Figure 4.1: Flow chart of recruitment and adherence



4.4.1.3 Missing data

The recruitment questionnaire had no missing data. However, there was difficulty obtaining some GP records and others obtained were incomplete, so the GP record questionnaire had at least one missing data field for 44/358, 22/205, and 46/302 individuals in positive follow-up, negative follow-up, and negative controls respectively. The highest number of missing fields was 13/16 seen in five participants (one positive follow-up and four negative follow-up), these individuals were all students who had incomplete notes sent from previous practices and were removed from analysis of GP records. More than 25% of data was missing for only two fields in the GP record questionnaire: time since most recent district nurse appointment and time since most recent surgery.

4.4.1.4 Adherence with follow-up

In the 18 months following 17 December 2008 34/360 (9%) individuals in positive follow-up were censored (24 because they missed more than three swabs; five because they withdrew from the study; three who moved out of area; and two for unrecorded reasons). In total, in positive follow-up in the first 18 months (including swabs from individuals who were censored), 2548 swabs were sent of which 2030 (80%) were returned, of these 1917 (94%) grew *S. aureus*. Of the 210 individuals in negative follow-up 21 (10%) were censored in the first 18 months (13 because they missed more than three swabs; five because they withdrew from the study; two who moved out of area; and one for unrecorded reasons). In total, for individuals in negative follow-up in the first 18 months (including swabs from individuals who were censored), 1165 swabs were sent, 986 (85%) were returned and 79 (8%) of these grew *S. aureus.* All censored individuals had samples included in analysis up to the date of censorship.

Continuation with, and compliance to follow-up was high, with the majority of participants returning all of their swabs, and good responses were received to telephone, text-message and email reminders about overdue swabs. Two-hundred and thirty-one of 260 (64%) participants in positive follow-up and 127/210 (60%) in negative follow-up returned every swab. The median number of swabs sent to individuals in positive follow-up was seven (lower quartile (LQ)=6, upper quartile (UQ)=8) and the median number received was also seven (LQ=5, UQ=8). For individuals in negative follow-up a

median of six swabs was sent (LQ=5, UQ=6) and five (LQ=4, UQ=6) received. Individuals in negative follow-up have so far returned fewer swabs, since recruitment into negative follow-up began nine months after positive follow-up. Swabs that were returned took a median of 11 days between being sent out and being received; time taken to return the swab did not correlate with whether the swab grew *S. aureus* (*P*=0.39).

4.4.2 Cross-sectional *S. aureus* carriage in the community

Three hundred and sixty of 1123 (32%) participants carried *S. aureus* at baseline. Nine (0.8% of individuals; 2.5% of *S. aureus* carriers) carried MRSA. Seventy-nine of 986 (8%) swabs and 41/210 (20%) individuals in negative follow-up had *S. aureus* growth from at least one swab.

4.4.3 Host environmental factors and cross-sectional carriage

Univariate analyses were used to identify host environmental factors correlating with *S. aureus* cross-sectional carriage. All factors from the personal details questionnaire and GP and hospital records were investigated using *S. aureus* swab test result at baseline as the outcome variable. Compared to non-carriers, participants with a swab positive for *S. aureus* at baseline were younger in age (*P*=0.048), more likely to be male (*P*<0.001), to currently be in employment (*P*=0.003), and to participate in contact sport (*P*=0.04) (Table 4.6). Odds ratios shown in table 4.6 suggest that for every year of age an individual's chance of carrying *S. aureus* decreases by approximately 1% and males are ~1.62 times more likely to carry *S. aureus* than females. Information from GP records did not contribute any further statistically significant risk factors for *S. aureus* colonisation (Table 4.7). Since only nine individuals carried MRSA risk factors correlated with meticillin resistance were not formally investigated, however their details are displayed in tables 4.6 and 4.7.

Table 4.6: Information from participant questionnaire

	Total (n=1123) N (%) or median (IQR)†	**MRSA (n=9) N (%) or median (IQR)†**	***S. aureus* carriage at baseline (n=360) N (%) or median (IQR)†**	**No carriage (n=763) N (%) or median (IQR)†**	**Univariate *P* value carriage versus no carriage**	**Univariate Odds Ratio (95% CI)**
Age†	55 (36;67)	64 (47;76)	53 (33;65)	56 (39;68)	0.048	0.99 (0.99;1.00)
Male	483 (43%)	4 (44%)	184 (51%)	299 (39%)	<0.001	1.62 (1.26;2.09)
Ethnic background:						
White British/Irish	1017 (91%)	9 (100%)	327 (91%)	690 (90%)	0.90	1.05 (0.68;1.61)
Other white background	64 (6%)	0 (0%)	21 (6%)	43 (6%)		1.04 (0.61;1.78)
Other ethnicity	42 (4%)	0 (0%)	12 (3%)	30 (4%)		0.84 (0.42;1.67)
Employment:						
Current	520 (46%)	3 (33%)	190 (53%)	330 (43%)	0.003	1.47 (1.14;1.89)
Healthcare related	257 (23%)	1 (11%)	76 (21%)	181 (24%)	0.36	0.86 (0.64;1.17)
Healthcare related in year prior to baseline	102 (9%)	0 (0%)	29 (8%)	73 (10%)	0.44	0.83 (0.53;1.30)
Number of household members:						
Lives alone	176 (16%)	3 (33%)	59 (16%)	117 (15%)	0.42	1.08 (0.77;1.52)
One other member	485 (43%)	4 (44%)	142 (39%)	343 (45%)		0.80 (0.62;1.03)
Two other members	194 (17%)	2 (22%)	70 (19%)	124 (16%)		1.24 (0.90;1.72)
Three other members	156 (14%)	0 (0%)	53 (15%)	103 (14%)		1.12 (0.77;1.58)
Four other members	63 (6%)	0 (0%)	24 (7%)	39 (5%)		1.33 (0.78;2.24)
Five or more other members	31 (3%)	0 (0%)	7 (2%)	24 (3%)		0.61 (0.26;1.43)
Shared accommodation with multiple members	18 (2%)	0 (0%)	5 (1%)	13 (2%)		0.81 (0.29;2.30)
Number of household members with healthcare contact:						
Zero	901 (80%)	7 (78%)	297 (83%)	604 (79%)	0.14	1.24 (0.90;1.71)
One	209 (19%)	2 (22%)	57 (16%)	152 (20%)		0.76 (0.54;1.06)
Two or more	13 (1%)	0 (0%)	6 (2%)	7 (1%)		1.83 (0.61;5.49)
Any household members with healthcare contact	222 (20%)	2 (22%)	63 (18%)	159 (21%)	0.20	0.81 (0.58;1.11)
Participation in contact sport	372 (33%)	3 (33%)	135 (38%)	237 (31%)	0.04	1.33 (1.02;1.73)
Looks after anyone with a disability	181 (16%)	5 (56%)	65 (18%)	116 (15%)	0.23	1.23 (0.88;1.71)
Looks after anyone with problems due to old age	211 (19%)	3 (33%)	60 (17%)	151 (20%)	0.22	0.81 (0.58;1.13)
Receives personal care, domestic help or other care	19 (2%)	0 (0%)	5 (1%)	14 (2%)	0.81	0.75 (0.27;2.11)

† data represented by median (IQR)

Table 4.7: Information from GP and hospital records

	Total (n=860) N (%) or median (IQR) †	MRSA (n=9) N (%) or median (IQR)†	*S. aureus* carriage at baseline (n=357) N (%) or median (IQR)†	No carriage (n=503) N (%) or median (IQR)†	Univariate *P* value carriage versus no carriage	Univariate Odds Ratio (95% CI)
In-patient visits:						
Ever *Unknown*	689 (80%) *10 (1%)*	8 (89%)	295 (83%) *4 (1%)*	394 (78%) *6 (1%)*	0.28	1.33 (0.93;1.90) *1.18 (0.32;4.37)*
Number in previous year† *(Unknown = 57)*	0 (0;0)	1 (0;1)	0 (0;0)	0 (0;0)	0.40	1.01 (0.97;1.05)
Days since most recent† *(Unknown = 171)*	1022 (313;2879)	342 (252;907)	986 (346;2853)	1046 (291;3169)	0.35	1.00 (1.00;1.00)
Out-patient appointments:						
Ever *Unknown*	801 (93%) *12 (1%)*	9 (100%)	330 (92%) *4 (1%)*	471 (94%) *8 (2%)*	0.54	0.73 (0.41;1.32) *0.52 (0.14;1.97)*
Number in previous year† *(Unknown = 86)*	1 (0;3)	2 (0;7)	1 (0;3)	1 (0;3)	0.52	1.01 (0.97;1.05)
Days since most recent† *(Unknown = 59)*	263 (66;963)	307 (85;658)	299 (90;912)	231 (52;984)	0.18	1.00 (1.00;1.00)
GP appointments:						
Ever *Unknown*	847 (98%) *8 (1%)*	9 (100%)	353 (99%) *1 (0.3%)*	494 (98%) *7 (1%)*	0.22	0.48 (0.08;2.87) *0.10 (0.01;1.50)*
Number in previous year† *(Unknown = 8)*	4 (2;7)	7 (3;11)	4 (2;7)	3 (2;6)	0.23	1.00 (0.98;1.04)
Days since most recent† *(Unknown = 13)*	67 (24;170)	47 (34;105)	62 (22;155)	72 (26;182)	0.30	1.00 (1.00;1.00)
Practice nurse appointments:						
Ever *Unknown*	824 (95%) *12 (1%)*	9 (100%)	345 (96%) *2 (1%)*	479 (94%) *10 (2%)*	0.25	1.01 (0.44;2.30) *0.28 (0.05;1.57)*
Number in previous year† *(Unknown = 12)*	2 (1;4)	3 (2;5)	2 (1;4)	2 (1;4)	0.52	0.98 (0.95;1.01)
Days since most recent† *(Unknown = 36)*	109 (32;320)	43 (13;53)	108 (29;318)	109 (34;326)	0.48	1.00 (1.00;1.00)
District nurse appointments:						
Ever *Unknown*	137 (16%) *16 (2%)*	3 (33%)	65 (18%) *4 (1%)*	72 (14%) *12 (2%)*	0.14	1.31 (0.91;1.90) *0.48 (0.15;1.52)*
Number in previous year† *(Unknown = 16)*	0 (0;0)	0 (0;0)	0 (0;0)	0 (0;0)	0.08	0.61 (0.35;1.08)
Years since most recent district nurse appointment:						
Never	70 (82%)	6 (67%)	289 (81%)	419 (83%)	0.10	0.84 (0.59;1.19)
0-1 years	31 (4%)	0 (0%)	8 (2%)	23 (5%)		0.48 (0.21;1.08)

	Total (n=860) N (%) or median (IQR) †	MRSA (n=9) N (%) or median (IQR)†	*S. aureus* carriage at baseline (n=357) N (%) or median (IQR)†	No carriage (n=503) N (%) or median (IQR)†	Univariate *P* value carriage versus no carriage	Univariate Odds Ratio (95% CI)
1-2 years	14 (2%)	1 (11%)	8 (2%)	6 (1%)		1.90 (0.66;5.52)
2-3 years	6 (1%)	0 (0%)	2 (1%)	4 (1%)		0.70 (0.13;3.86)
3-4 years	7 (1%)	0 (0%)	5 (1%)	2 (0.4%)		3.56 (0.69;18.4)
4-5 years	3 (0.4%)	0 (0%)	2 (1%)	1 (0.2%)		2.83 (0.26;31.3)
5-10 years	27 (3%)	1 (11%)	15 (4%)	12 (2%)		1.79 (0.83;3.88)
10-20 years	27 (3%)	1 (11%)	13 (4%)	14 (3%)		1.32 (0.61;2.84)
>20 years	22 (2%)	0 (0%)	12 (3%)	10 (2%)		1.71 (0.73;4.01)
Unknown	16 (2%)	0 (0%)	4 (1%)	12 (2%)		0.46 (0.15;1.45)
Co-morbidities:						
Has a long-term illness *Unknown*	464 (54%) *9 (1%)*	7 (78%)	202 (56%) *2 (1%)*	262 (52%) *7 (1%)*	0.27	1.18 (0.90;1.55) *0.44 (0.09;2.13)*
Has had chemotherapy *Unknown*	28 (3%) *25 (3%)*	1 (11%)	(2%) *11 (3%)*	21 (4%) *14 (3%)*	0.19	0.46 (0.19;1.09) *1.08 (0.49;2.42)*
Has had renal dialysis *Unknown*	4 (0.5%) *24 (3%)*	1 (11%)	3 (1%) *10 (3%)*	1 (0.2%) *14 (3%)*	0.49	4.26 (0.44;41.1) *1.01 (0.44;2.31)*
Has had surgery *Unknown*	632 (73%) *16 (2%)*	9 (100%)	266 (74%) *8 (2%)*	366 (72%) *8 (2%)*	0.57	1.13 (0.82;1.56) *1.55 (0.56;4.30)*
Years since most recent surgery:						
Never	212 (25%)	0 (0%)	83 (23%)	129 (25%)	0.55	0.88 (0.64;1.21)
0-1 years	101 (12%)	1 (11%)	35 (10%)	66 (13%)		0.72 (0.47;1.12)
1-3 years	122 (14%)	1 (11%)	47 (13%)	75 (15%)		0.87 (0.58;0.28)
3-5 years	61 (7%)	1 (11%)	28 (8%)	33 (7%)		1.21 (0.72;2.04)
5-10 years	112 (13%)	3 (33%)	48 (13%)	64 (13%)		1.07 (0.71;1.60)
10-20 years	131 (15%)	1 (11%)	58 (16%)	73 (14%)		1.14 (0.79;1.66)
>20 years	105 (12%)	2 (22%)	50 (14%)	55 (11%)		1.33 (0.88;2.00)
Unknown	16 (2%)	0 (0%)	8 (2%)	8 (2%)		1.42 (0.53;3.82)
Has been prescribed oral steroids *Unknown*	133 (15%) *43 (5%)*	1 (11%)	51 (14%) *15 (4%)*	82 (16%) *28 (6%)*	0.45	0.84 (0.57;1.23) *0.72 (0.38;1.38)*
Treatment for skin conditions in past month *Unknown*	97 (11%) *18 (2%)*	2 (22%)	50 (14%) *7 (2%)*	47 (9%) *11 (2%)*	0.11	1.58 (1.03;2.41) *0.94 (0.36;2.46)*
Vascular access in past month *Unknown*	560 (65%) *38 (4%)*	2 (22%)	240 (67%) *16 (4%)*	320 (64%) *22 (4%)*	0.50	1.20 (0.89;1.61) *1.16 (0.58;0.23)*
Has had a catheter *Unknown*	208 (24%) *44 (5%)*	4 (44%)	79 (22%) *18 (5%)*	129 (26%) *26 (5%)*	0.49	0.82 (0.59;1.13) *0.93 (0.50;1.73)*
Antibiotics:						
Ever *Unknown*	809 (94%) *28 (3%)*	9 (100%)	333 (93%) *11 (3%)*	476 (95%) *17 (3%)*	0.34	0.54 (0.23;1.24) *0.50 (0.16;1.53)*
Number in previous year† *(Unknown = 28)*	0 (0;1)	1 (0;2)	0 (0;1)	0 (0;1)	0.44	0.96 (0.88;1.05)
Days since most recent† *(Unknown = 51)*	631 (179;2043)	131 (105;241)	542 (183;1863)	672 (170;2085)	0.60	1.00 (1.00;1.00)

	Total (n=860) N (%) or median (IQR) †	**MRSA (n=9) N (%) or median (IQR)†**	***S. aureus* carriage at baseline (n=357) N (%) or median (IQR)†**	**No carriage (n=503) N (%) or median (IQR)†**	**Univariate *P* value carriage versus no carriage**	**Univariate Odds Ratio (95% CI)**
Had MRSA previously	7 (1%)	0 (0%)	4 (1%)	3 (1%)	0.48	1.86 (0.41;8.36)
Unknown	*46 (5%)*		*16 (4%)*	*30 (6%)*		*0.74 (0.40;1.39)*
Had MSSA previously	48 (6%)	1 (11%)	20 (6%)	28 (6%)	0.66	0.99 (0.55; 1.79)
Unknown	*46 (5%)*		*16 (4%)*	*30 (6%)*		*0.74 (0.40;1.38)*

† data represented by median (IQR)

While univariate analyses provide an informative assessment of risk factors, it was possible that some variables included in the univariate model were correlated, perhaps obscuring the true effects. Multivariate analyses were undertaken separately on information from the personal details questionnaire and GP and electronic records (the two were separated because the latter was collected for fewer participants). All data from the participant questionnaire were included in the regression model, except for when there were dependencies between variables, for example whether the participant lived alone and the number of household members, in which case only one measure was used (in this example, number of household members, as it contained more information). All of the variables had a linear relationship with *S. aureus* carriage so one degree of freedom was used. Variables left in the model at $P<0.05$ were male sex ($P<0.001$); current employment ($P=0.002$); participation in contact sport ($P=0.03$); and looking after someone with a disability ($P=0.02$) all which increased the risk of *S. aureus* carriage (Table 4.8). Results from the multivariate model were largely consistent with the univariate analysis (Table 4.6) in which male sex, current employment and sport were all significant risk factors for *S. aureus* carriage, with similar odds ratios. In the univariate analysis looking after someone with a disability was not a significant predictor of *S. aureus* carriage, although it did tend to increase risk. Additionally, age was significant in the univariate analysis, but not included in the final multivariate model, probably due to its correlation with other factors included in the multivariate model, i.e. participation in contact sport and current employment.

Table 4.8: Multivariate model for the effect of factors identified from the participant questionnaire on carriage of *S. aureus* at baseline

	Odds Ratio	LCI	UCI	*P* value
Male	1.68	1.30	2.17	<0.001
Current employment	1.48	1.15	1.92	0.002
Participation in contact sport	1.34	1.02	1.74	0.03
Looks after someone with a disability	1.57	1.07	2.30	0.02

Data from GP records was handled in a similar way, excluding dependent variables, for example whether the participant had had a GP appointment, number of GP appointments in previous year and days since most recent GP appointment. In these circumstances variables containing information on days or years since events were included, since they were the most informative. Long-term illness was not included as this was indicated by other healthcare-related variables, and data on previous MRSA and MSSA was combined as they had identical missing values. The relationship between *S. aureus* carriage and days between baseline and most recent out-patient appointment was non-linear; further investigation revealed a relationship with two linear parts, changing at approximately 30 days (Figure 4.2), suggesting an immediate but short term protective effect of a recent out-patient appointment. Therefore days since outpatient appointment was categorised into whether the participant had been an out-patient in the last 30 days or not (models including days since out-patient appointment and out-patient within 30 days were equivalent by AIC). Variables found to be significant at $P<0.05$ were treatment for skin conditions in the month prior to baseline ($P=0.04$), which increased risk of *S. aureus* carriage and out-patient appointment within 30 days ($P=0.01$), which decreased risk of *S. aureus* carriage (Table 4.9). Neither risk factors were significant in univariate analysis, highlighting the increased power of a multivariate approach compared to univariate analysis.

Figure 4.2: Relationship between time since most recent out-patient appointment and *S. aureus* carriage

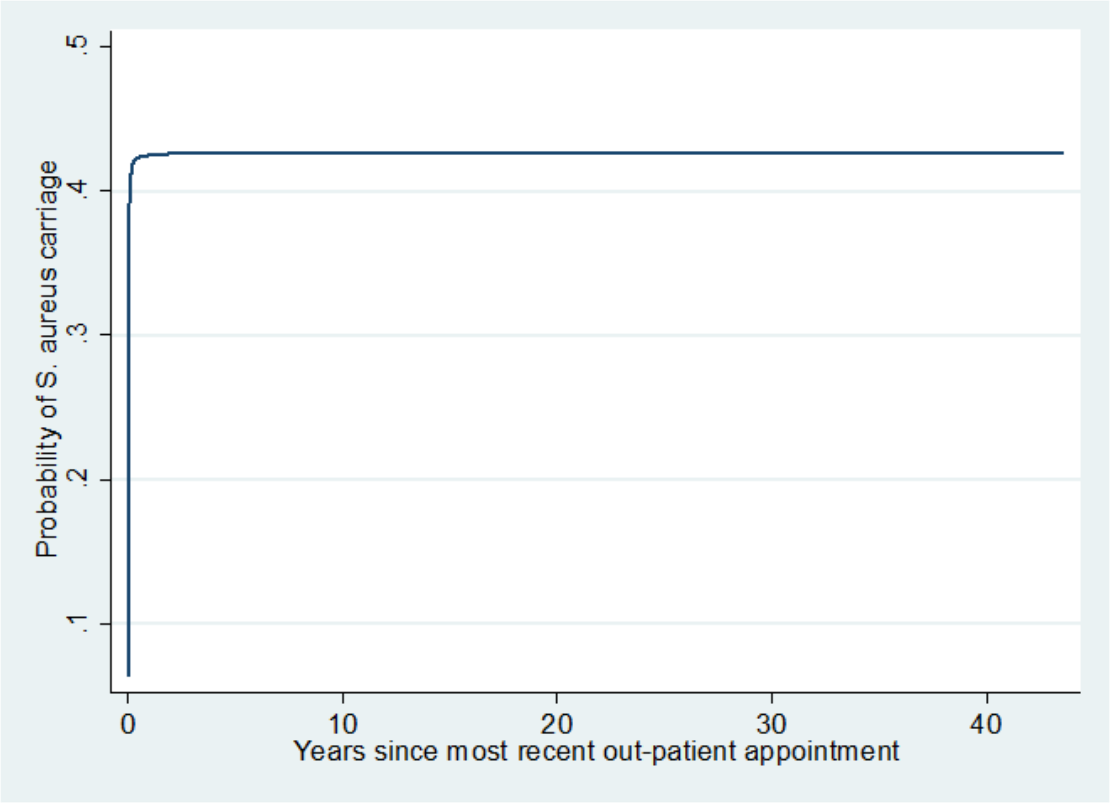


Table 4.9: Multivariate model for the effect of factors identified from GP records on carriage of *S. aureus* at baseline

	Odds Ratio	LCI	UCI	*P* value
Received treatment for skin conditions in past month	1.56	1.03	2.43	0.04
Unknown	*0.98*	*0.96*	*0.37*	*2.57*
Out-patient appointment in previous 30 days	0.57	0.37	0.89	0.01

Multivariate analysis identified six factors correlated with increased risk of a participant having cross-sectional *S. aureus* carriage (Table 4.8 and 4.9). The strongest effects were whether the participant was male (OR=1.68); whether the participant was in current employment (OR=1.48); and whether the participant had had an out-patient appointment in the past 30 days (OR=0.37). Interestingly the odds ratio showed that a recent out-patient appointment was protective against *S. aureus* carriage. Out-patient appointments are a marker for illness, so perhaps this protective effect could be explained by

prescription of antibiotics. However, the same effect was not seen when directly investigating antibiotics, so it is possible that other unknown factors are also involved. Days since out-patient appointment was not significant in the univariate model, however the *P* value for whether the participant had been an out-patient in the past 30 days was highly significant (*P*=0.01) suggesting a strong effect on *S. aureus* carriage. Participation in contact sport was significant in both the univariate and multivariate analysis, suggesting this is a reliable predictor of *S. aureus* carriage. Looking after someone with a disability and skin conditions were not significant in the univariate analysis nor were they significant at *P*<0.01 in the multivariate analysis and so are less likely to have an important effect on *S. aureus* carriage.

4.4.3.1 Cross-sectional MRSA carriage

Two of nine individuals with MRSA carriage fitted the CDC definition for CA-MRSA. However, they were both persistent carriers of healthcare-associated t032 MRSA and had GP and practice nurse appointments in the year prior to baseline. Therefore, these individuals were unlikely to have acquired MRSA in hospital, however may have acquired hospital associated strains during other healthcare contact.

4.4.3.2 Host environmental factors of individuals in negative follow-up

Participants in negative follow-up comprised two groups; participants recruited from Bampton Medical Practice (n=128) and students recruited in University Freshers week (n=82). Students were significantly more likely to have *S. aureus* from at least one follow-up swab with 28 (34%) acquiring *S. aureus*, as opposed to 13 (10%) of participants from Bampton (*P*<0.001).

Univariate analysis of information from the participant questionnaire showed that compared to individuals in negative follow-up who did not have *S. aureus* identified from any sample, individuals in negative follow-up with *S. aureus* were younger, with a median age of 19 compared to 44 years (*P*<0.001); and had more household members (*P*=0.001). This information was entered into a multivariate model with *S. aureus* growth as the dependent variable, from which age was the only significant predictor of *S. aureus* growth (OR=0.95, LCI=0.93, UCI=0.97, *P*<0.001).

To ensure the relationship between younger age and *S. aureus* growth in negative follow-up was not caused by the students, univariate analysis was performed on only participants recruited from Bampton Medical Practice. Younger age remained a significant predictor of *S. aureus* identification in individuals in negative follow-up from Bampton, those with *S. aureus* had a median age of 39 years compared to 59 for individuals who did not (*P*=0.02). Additionally, when student status was included in the multivariate model, the relationship with age remained. Information from GP and hospital records was entered into a multivariate model with *S. aureus* growth as the dependent variable, which also included age, as age confounds many healthcare risk factors. Because data was missing from many students' GP records, leading to co-linearity in the model, 16 further participants had to be removed before running the model. Again, the multivariate model revealed only younger age to be associated with detection of *S. aureus* in individuals in negative follow-up (*P*=0.002).

4.4.4 Longitudinal carriage

4.4.4.1 False negative swabs

Before investigating longitudinal carriage it was necessary to determine the rate of false negative swabs, since classifying false negatives, whereby a participant was carrying *S. aureus* that was not detected by the swab, as true negatives could change inferences about carriage persistence and duration. False negative swabs might be caused by limitations in the sensitivity of nasal swabbing combined with fluctuations in the amount of *S. aureus* carried. Additionally, the use of self-swabbing, required for the study protocol, may have increased chances of false negative swabs if participants did not swab their noses sufficiently or consistently.

Only individuals in positive follow-up were used to estimate false negative rates, since only 41 individuals in negative follow-up had a positive swab for *S. aureus* of which 75% returned two swabs or fewer. Analysis considered false negatives at the level of the individual. All combinations of positive and negative swabs were calculated for up to three swabs in a row, and the effect of these previous swabs on whether the current swab was positive was investigated (Table 4.10). Having two or three previous consecutive negative or positive swabs was highly predictive of a subsequent swab being negative or positive respectively. However, having just one previous negative swab preceded

by a positive swab was not predictive of the subsequent swab being negative. This indicated that one single negative swab is likely to be a false negative, whereas two negative swabs are likely to indicate loss of carriage.

Table 4.10: Number and percentage of positive and negative swabs following combinations of up to three previous swab results

Previous swab results	Number followed by + swab	Percent followed by + swab	Number followed by – swab	Percent followed by – swab
-	44	16%	233	84%
+	1513	92%	139	8%
--	19	11%	149	89%
+-	25	23%	86	77%
++	1195	94%	73	6%
-+	23	62%	14	38%
---	13	14%	82	86%
+--	6	8%	65	92%
-+-	2	20%	8	80%
++-	15	29%	37	71%
+++	913	95%	49	5%
-++	14	74%	5	26%
+-+	13	59%	9	41%
--+	10	67%	5	33%

Logistic regression, with the result of the current swab as the dependent variable, was used to test the inference that a single negative swab is a false negative. The regression considered the effect of the two previous swabs on the current swab, including an interaction between them. It showed that individually the two previous swabs were both predictive of the current swab, with an additional significant interaction between them (P=0.001), indicating that two previous positive swabs together were even more predictive of the current swab also being positive for the same *spa* type (OR=123). Addition of a third previous swab to the regression model did not have a significant effect on the current swab (P=0.5).

Similar analysis was performed at the level of the strain, whereby results were separated by *spa* type and subsequent swabs were considered positive if they all shared the same *spa* type. Results at the strain level were consistent with analysis at the level of the individual, and all results were supported by previous research (Nouwen, Ott et al. 2004). It was therefore decided that two subsequent negative swabs would be regarded as strain loss, whereas a single negative nasal swab (followed by

a positive) was a false negative, and ignored in analysis. The false negative rate was used when classifying *S. aureus* longitudinal carriage.

4.4.4.2 Investigation of longitudinal carriage by carriage group

A four-group classification system was designed to investigate longitudinal carriage, whilst accounting for changes in the type of carried bacteria and the possibility of co-colonisation. It was developed considering individuals in positive follow-up only, however it could also be used to interpret the carriage behaviour of *S. aureus* in individuals in negative follow-up.

The first two carriage groups comprised individuals who carried only one *spa* type throughout in order to identify participants who did not appear to acquire new strains of *S. aureus*. These were (i) single-persistent, designed to identify people always colonised by the same strain, where the same single *spa* type was found on all swabs without two successive negative swabs (Figure 4.3); and (ii) single-intermittent, designed to identify participants re-acquiring the same strain, where the same single *spa* type was found on all returned swabs and at least two successive swabs failed to grow *S. aureus* (Figure 4.4). The second two groups included individuals with more than one *spa* type. Group (iii) multiple-persistent was designed to encompass individuals who were persistent carriers of one strain, but also had additional *spa* types. It included all individuals who had a single-persistent *spa* type (not missing two in a row) as well as having other *spa* types, which could represent new acquisitions (Figure 4.5). Group (iv) multiple-changing comprised people with multiple *spa* types, with no consistency, so the *spa* type appeared to be changing over time, this included any individuals with three or more swabs returned who did not fit into any of the above groups (Figure 4.6). Additionally, group (v) unclassifiable was individuals who returned less than four swabs.

Figure 4.3: Two examples of single-persistent carriers

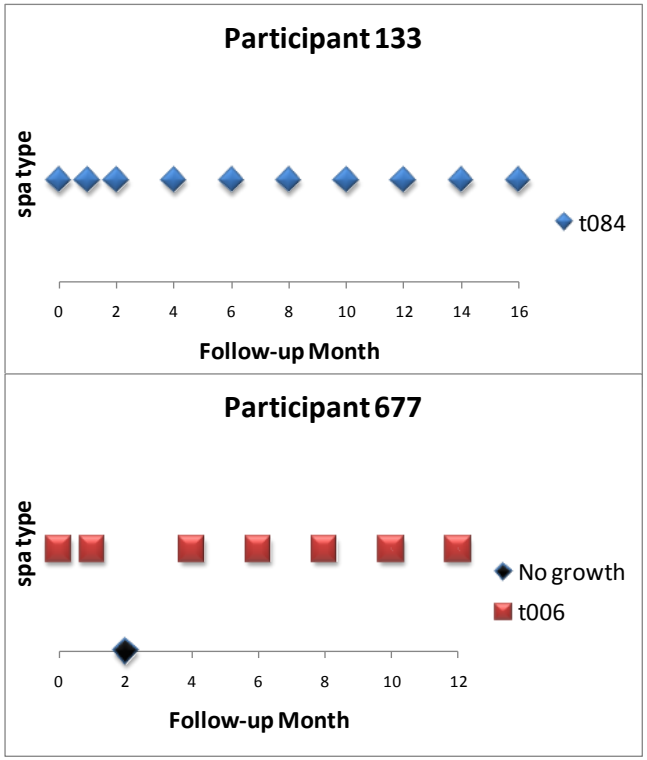


Figure 4.4: Two examples of single-intermittent carriers

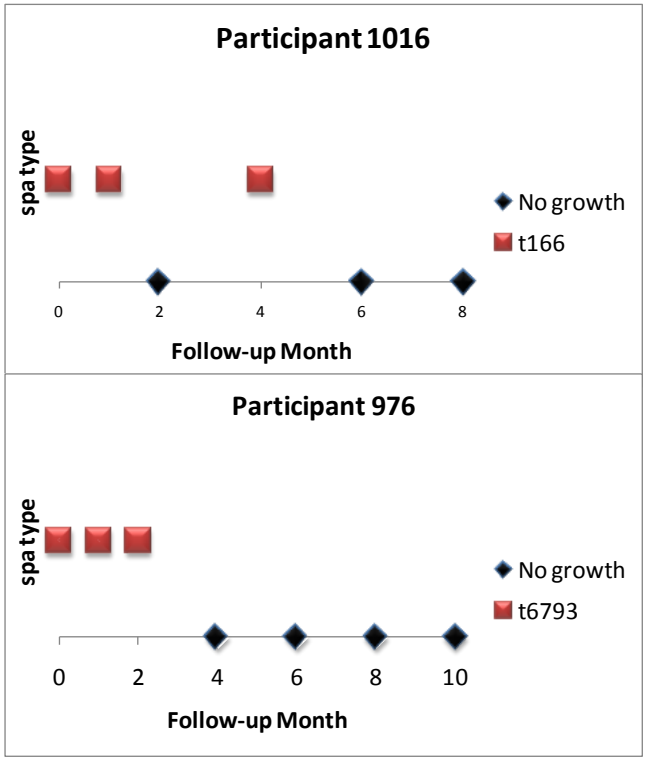


Figure 4.5: Three examples of multiple-persistent carriers

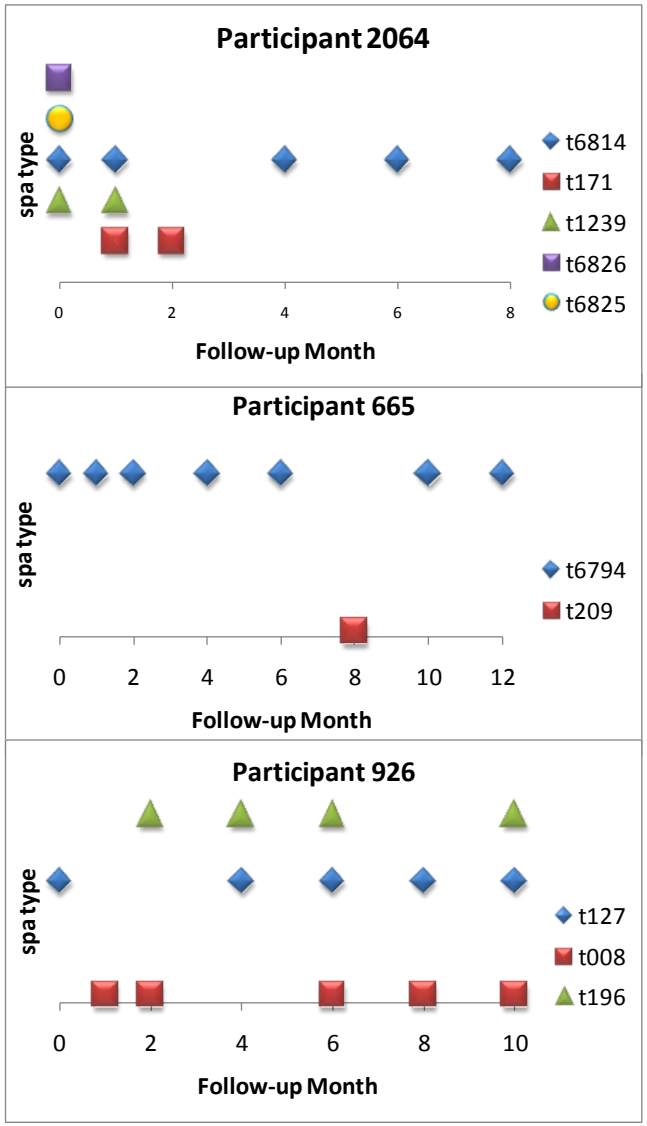
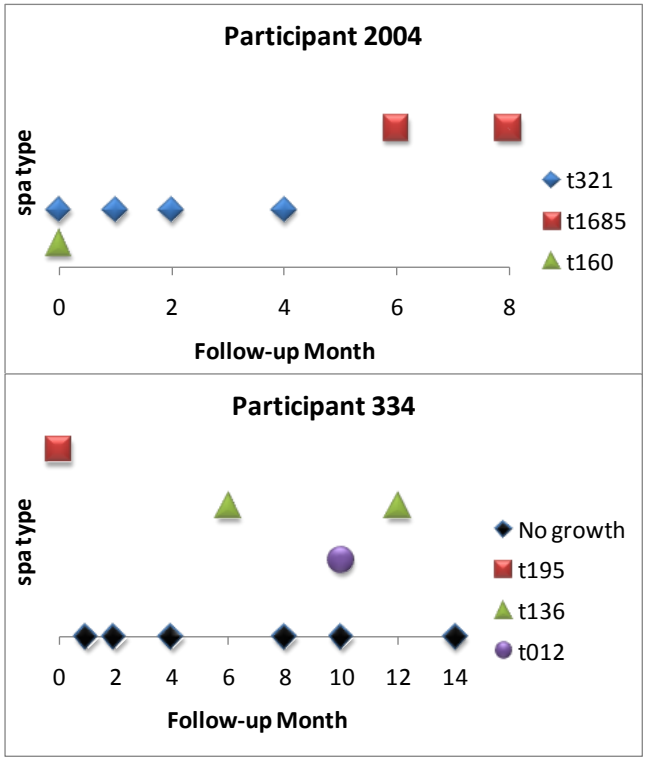


Figure 4.6: Two examples of multiple-changing carriers



For individuals in positive follow-up, the most common carriage group was single-persistent, consisting of 47% of carriers, the other groups were distributed more evenly (Table 4.11). The number of swabs returned did not differ significantly between individuals in each group (Table 4.11).

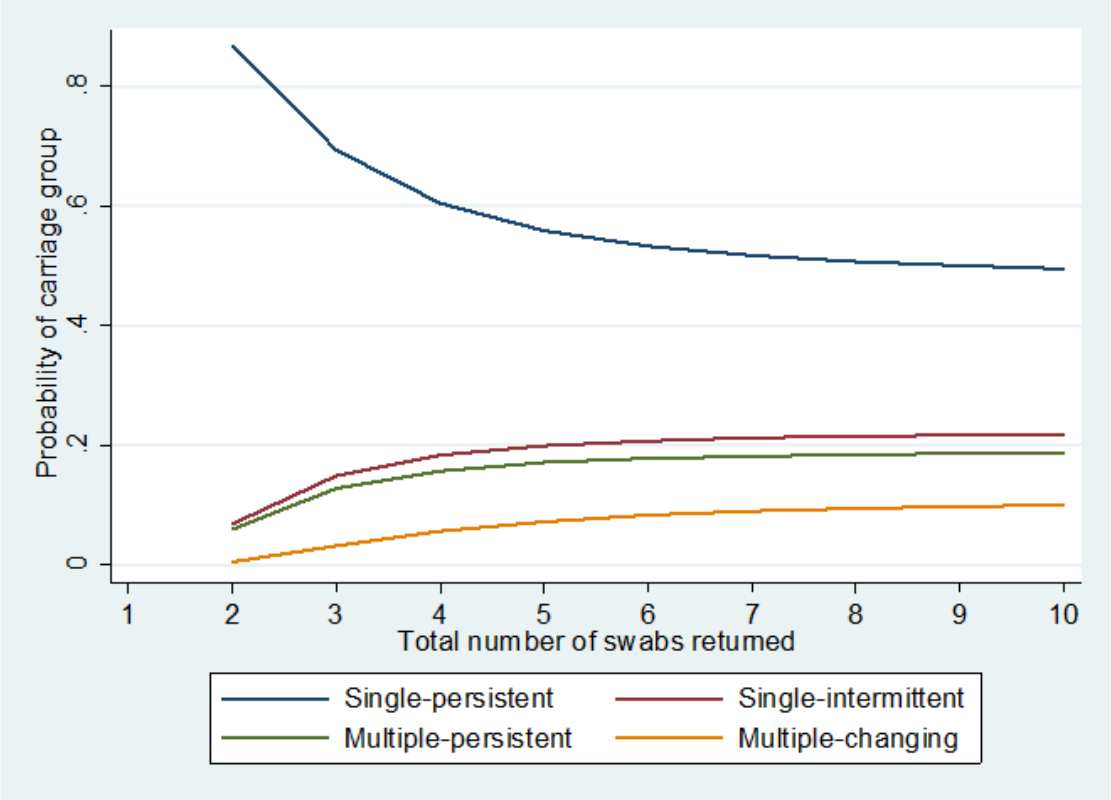
Individuals in negative follow-up were also classified into carriage groups and were distributed significantly differently between the groups from individuals in positive follow-up (P<0.001), with higher proportions of single-intermittent carriers and multiple-changing carriers (Table 4.11). This suggests that individuals in negative follow-up may represent a different group of individuals who are more likely to be transient carriers.

Table 4.11: Carriage groups for individuals in positive follow-up *(negative follow-up represented in parentheses)*

Carriage group	**Number of participants in positive follow-up *(negative follow-up)***	**Details of number of swabs returned by participant**				
		Median	**Lower Quartile**	**Upper Quartile**	**Maximum**	**Minimum**
Single-persistent	170 *(4)*	7 *(5)*	6 *(5)*	8 *(5)*	10 *(5)*	4 *(5)*
Single-intermittent	67 *(22)*	7*(6)*	6 *(5)*	9 *(6)*	10 *(6)*	4 *(4)*
Multiple-persistent	59 *(2)*	7 *(5.5)*	5 *(5)*	8 *(6)*	10 *(6)*	4 *(5)*
Multiple-changing	28 *(8)*	6 *(5)*	6 *(5)*	8 *(6)*	10 *(6)*	5 *(5)*
Unclassified	36 *(5)*	2 *(*3)	1 *(3)*	3 *(3)*	3 *(3)*	1 *(3)*

In order to check the robustness of classification and to determine the minimum number of returned swabs required to classify individuals into carriage groups, logistic regression was carried out investigating the non-linear effect of the number of swabs returned on carriage group. Only individuals from positive follow-up were considered, since individuals in negative follow-up had returned fewer swabs, and were less likely to be persistent carriers, which confounded results. Number of swabs returned was limited to greater than one in the regression as when only one swab was returned it was impossible to distinguish between carriage groups. The resulting regression curve shows that by excluding individuals with less than four swabs the relationship between number of swabs and carriage group became close to linear (Figure 4.7) whilst only necessitating exclusion of 36/360 (10%) of participants in positive follow-up.

Figure 4.7: Effect of number of swabs returned on assignment into carriage group



In order to gain maximum use of the resolution produced by *spa* typing, carriage groups were defined using individual *spa* types. However, a cost of this was that the groups did not incorporate information regarding similarity of *spa* types. Single-persistent and single-intermittent carriers represented a conservative approach with respect to strain evolution, defining people as carriers of single strains only if they carried exactly the same *spa* type throughout. Despite this, 66% of individuals in negative follow-up were classified as carriers of a single strain.

Changing carriage group classification to be defined by *spa* CC would result in reclassification of seven individuals with multiple-changing carriage to single-persistent (n=5), multiple-persistent (n=1) and single-intermittent (n=1) carriers and 17 individuals with multiple-persistent carriage to single-persistent. Classification of carriage groups using CC would therefore result in 192 individuals in positive follow-up classified as single-persistent and 68 single-intermittent, so 72% of individuals in positive follow-up would be regarded as carrying a single strain, an increase of 6%. For individuals in negative follow-up only one multiple-changing carrier became a multiple-persistent carrier.

Defining carriage groups represented a compromise between overestimation of evolution and overestimation of acquisition. Categorisation into carriage groups using *spa* CC would tend to overestimate strain change by evolution within the nose, particularly for the largest *spa* CC, CC012, which contained 64 *spa* types in this study, some of which were relatively distantly related. This is highlighted by individual 993 who carried two MRSA and one MSSA strain all in *spa* CC012 (Table 4.12). Classification by *spa* CC would class participant 993 as a single-persistent carrier, even though it is highly improbable that the MSSA strain t748 arose by evolution in the nose from the other MRSA strains identified in this participant's nose. However classification according to CC may overestimate acquisition, for example participant 132 was classified as a multiple-persistent carrier according to *spa* type but carried three *spa* types with highly similar repeat patterns (Table 4.12).

There is limited research investigating *S. aureus* carriage and evolution of *spa* types over time; one study suggests low levels of evolution within the nose with only 13/319 (4%) individuals deemed to have *spa* types that evolved within the nose (Boye and Westh 2011). Thus, since the likelihood of *S. aureus* evolution within the nose during this study is unknown; and because intermittent carriers with different *spa* types in the same CC were unlikely to represent evolution in the nose, it was decided to define carriage group using individual *spa* types. However, it must be kept in mind that multiple-persistent and multiple-changing carriers may contain up to 17/59 (29%) and 7/28 (25%) individuals respectively who have *spa* types that changed by evolution rather than acquisition.

Table 4.12: Examples of multiple *spa* types in the same BURP CC from two participants

Participant ID	BURP CC	*spa* type	Meticillin resistance	Repeat pattern
993	012	t012	Resistant	15-12-16-02-16-02-25-17-24-24
		t021	Resistant	15-12-16-02-16-02-25-17-24
		t748	Sensitive	15-12- 17-24
132	084	t228	Sensitive	14-12-34-34- 12-12-23-02-12-23
		t1885	Sensitive	14-12-34- 12-12-23-02-12-23
		t3097	Sensitive	14-12-34-34-34-12-12-23-02-12-23

4.4.4.3 Host environmental factors and carriage group

Epidemiological covariates may explain the existence of different carriage groups. To investigate the effect of epidemiology on carriage group only individuals in positive follow-up were considered, since individuals in negative follow-up were a small population with significantly different carriage behaviour (Table 4.11). Univariate analysis on factors from the participant questionnaire found only age to differ

between carriage groups (*P*=0.02), since older participants were more likely to be single-persistent carriers (Table 4.13, Figure 4.8). MFP regression showed evidence for significant non-linearity in the relationship between age and carriage type (Figure 4.8), therefore age was log transformed using a twofold reduction to aid interpretation by making the relationship linear (models including age and log transformed age were equivalent by AIC). Results from multivariate analysis supported those from univariate analysis and left only age in the model. Here, individuals with single-intermittent *spa* types and multiple-changing *spa* types were significantly younger than individuals with a single-persistent *spa* type (*P*=0.01 RRR=0.59 and *P*=0.02 RRR=0.52 respectively) (Table 4.14). Relative risk ratio (RRR) is similar to an odds ratio, but used when the dependent variable is categorical with more than two categories. Here the RRR represents the ratio of the probability of being in one carriage group for every increasing unit of age compared to a particular carriage group assigned as a reference group. Since age was log transformed with a twofold reduction, RRR of 0.59 indicates that for every increase in age unit, equivalent to doubling of age in years, a participant is 0.59 times as likely to have a single-intermittent *spa* type compared to the reference: single-persistent *spa* type.

Table 4.13: Information from the participant questionnaire separated by carriage group

	Single-persistent (n=170) N (%) or median (IQR)†	**Single-intermittent (n=67) or median (IQR)†**	**Multiple-persistent (n=59) or median (IQR)†**	**Multiple-changing (n=28) or median (IQR)†**	**Univariate *P* value**
Age†	56 (41;67)	48 (25;67)	50 (28;66)	51 (22;62)	0.02
Male	87 (51%)	28 (42%)	33 (56%)	16 (57%)	0.36
Ethnic background:					
White British/Irish	158 (93%)	59 (88%)	57 (97%)	24 (86%)	0.68
Other white background	9 (5%)	5 (7%)	2 (3%)	2 (7%)	
Other ethnicity	3 (2%)	3 (4%)	0 (0%)	2 (7%)	
Employment:					
Current	88 (52%)	38 (57%)	29 (49%)	14 (50%)	0.84
Healthcare related	36 (21%)	16 (24%)	10 (17%)	7 (25%)	0.76
Healthcare related in year prior to baseline	11 (6%)	11 (6%)	6 (9%)	4 (14%)	0.51
Number of household members:					
Lives alone	30 (18%)	12 (18%)	9 (15%)	3 (11%)	0.80
One other member	77 (45%)	21 (31%)	25 (42%)	13 (46%)	
Two or more members	61 (36%)	33 (49%)	24 (41%)	12 (43%)	
Shared accommodation with multiple members	2 (1%)	1 (1%)	1 (1%)	0 (0%)	
Has household member with healthcare contact	30 (17%)	17 (25%)	5 (8%)	4 (14%)	0.11
Participation in contact sport	53 (31%)	30 (45%)	24 (41%)	13 (46%)	0.13
Looks after anyone with a disability or old age	48 (28%)	16 (24%)	13 (22%)	5 (18%)	0.58
Receives personal care, domestic help or other care	4 (2%)	0 (0%)	1 (2%)	0 (0%)	0.99

† data represented by median (IQR)

Figure 4.8: Probability of assignment into each carriage group by age

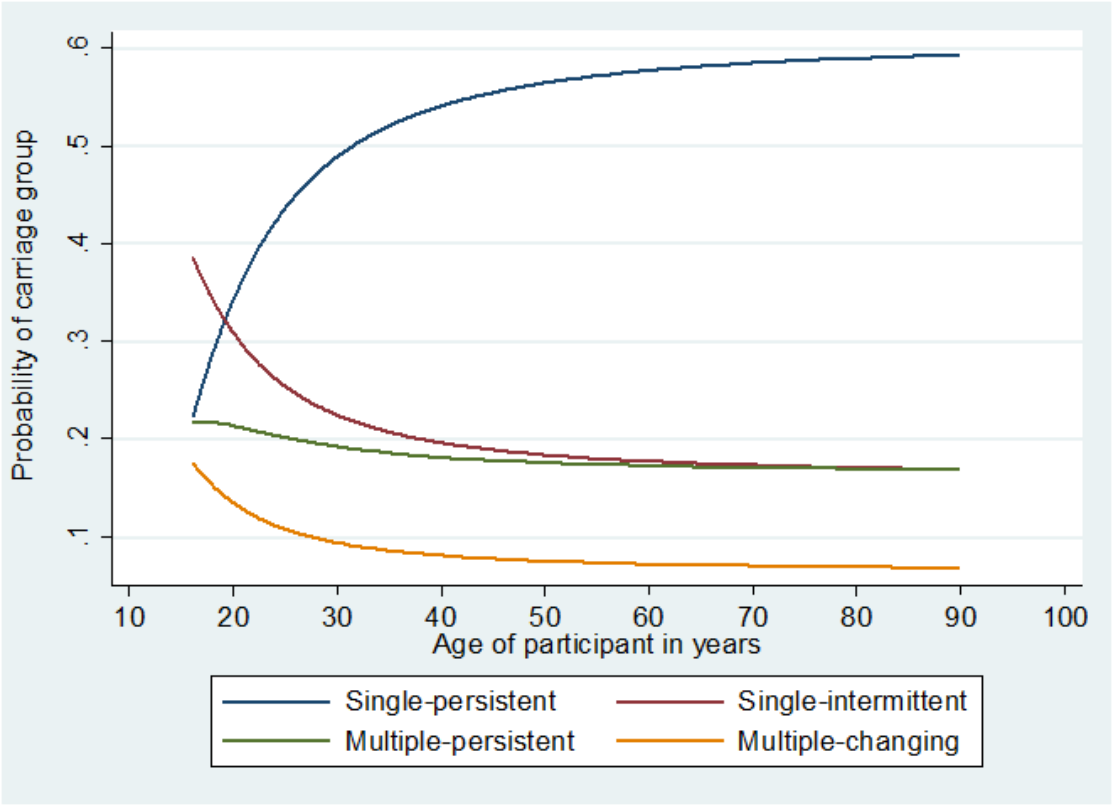


Table 4.14: Multivariate model for the effect of factors identified from the participant questionnaire on carriage group

	RRR	LCI	UCI	*P* value
Age:				
Single-intermittent versus single-persistent	0.59	0.39	0.90	0.01
Multiple-persistent versus single-persistent	0.69	0.43	1.03	0.07
Multiple-changing versus single-persistent	0.52	0.39	0.91	0.02

Univariate analysis of information from GP and hospital records showed that no factors differed between carriage groups (Table 4.15). (For conciseness some variables seen in tables 4.6 and 4.7 are not shown as they were epidemiologically unlikely to be related to carriage group, however all variables were tested and variables not shown did not significantly differ between carriage groups.) In table 4.15, instances of healthcare exposure in the year prior to baseline was chosen to represent healthcare exposure, since increased healthcare exposure in the year prior to baseline suggests continued increased exposure during the study period. Multivariate analysis of information from GP records used days since most recent healthcare contact rather than instances of healthcare contact due to further co-linearity and small numbers; and chemotherapy and renal dialysis were merged into

a single variable. Due to further co-linearity time since district nurse visit, skin break in the month prior to baseline and previous MSSA were also not included in the multivariate analysis. Variables left in the multivariate model with $P<0.05$ were: having had chemotherapy or renal dialysis ($P=0.02$, RRR=10.2) for multiple-changing *spa* compared to single-persistent *spa* type and days since out-patient appointment ($P=0.045$, RRR=1.00) for single-intermittent *spa* compared to single-persistent. Both of these had non-linear relationships with the dependent variable. However, on further investigation it was unlikely that either of these were reliable predictors of carriage group, since neither factor was significant in the univariate analysis and *P* values were not significant at the 0.1 level. Additionally, results from regression models using either chemotherapy or renal dialysis separately included neither factor in the final model, indicating that individually chemotherapy and renal dialysis were not associated with carriage group. This suggests that no information collected from GP or hospital records differs significantly between carriage groups, however perhaps with larger numbers some associations suggested here may become significant.

Table 4.15: Information from GP and hospital records separated by carriage group

	Single-persistent (n=169) N (%) or median (IQR)†	Single-intermittent (n=66) or median (IQR)†	Multiple-persistent (n=58) or median (IQR)†	Multiple-changing (n=28) or median (IQR)†	Univariate *P* value
Healthcare exposure in year prior to baseline:					
Number in-patient appointments† *(Unknown = 17)*	0 (0;0)	0 (0;0)	0 (0;1)	0 (0;1)	0.23
Number out-patient appointments† *(Unknown = 29)*	1 (0;2)	1(0;3)	1(0;3)	1 (0;3)	0.31
Number GP appointments† *(Unknown =1)*	4 (2;7)	4 (2;6)	5 (1;7)	4 (2;7)	0.91
Number practice nurse appointments† *(Unknown = 2)*	2 (1;4)	3 (1;5)	2 (1;5)	1 (1;3)	0.28
Number district nurse appointments† *(Unknown = 4)*	0 (0;0)	0 (0;0)	0 (0;0)	0 (0;0)	0.86
Surgery *(Unknown = 6)*	13 (8%)	7 (11%)	10 (17%)	2 (7%)	0.36
Other healthcare risk factors:					
Has a long-term illness *(Unknown=2)*	97 (57%)	37 (56%)	35 (60%)	17 (61%)	0.99
Has had chemotherapy *(Unknown=7)*	3 (2%)	4 (6%)	0 (0%)	0 (0%)	0.47
Has had renal dialysis *(Unknown=6)*	3 (2%)	3 (5%)	0 (0%)	0 (0%)	0.94
Ever been prescribed oral steroids *(Unknown=11)*	22 (13%)	11 (17%)	14 (24%)	0 (0%)	0.70
Treatment for skin conditions in past month *(Unknown=7)*	25 (15%)	7 (11%)	10 (17%)	2 (7%)	0.82
Vascular access in past month *(Unknown=16)*	108 (64%)	46 (70%)	41 (71%)	17 (61%)	0.82
Has had a catheter *(Unknown=18)*	34 (20%)	16 (24%)	18 (31%)	4 (14%)	0.54
Number antibiotics in year prior to baseline† *(Unknown = 10)*	0 (0;1)	0 (0;1)	0 (0;1)	0 (0;2)	0.41
Had MRSA previously *(Unknown = 11)*	1 (1%)	0 (0%)	2 (3%)	1 (4%)	0.96
Had MSSA previously *(Unknown=11)*	9 (5%)	6 (9%)	4 (7%)	1 (4%)	0.88

† data represented by median (IQR)

Comparison between carriage groups classified by individual *spa* types (Tables 4.14 and 4.15) and altered classification by *spa* CC revealed no difference in significant predictors from univariate and multivariate analysis of information from the personal questionnaire and GP records. However, the effect of age on carriage group was more significant, with a univariate *P* value of 0.01. Additionally, a

multivariate model identical to that described in section 5.10.1, but with carriage group classified by *spa* CC as the dependent variable, found that individuals with single-persistent carriage were significantly older than individuals in every other group, including multiple-persistent, which was not significantly different when carriage groups were defined by *spa* type (*P*=0.01 RRR=0.60 for single-intermittent; *P*=0.02 RRR=0.58 for multiple-persistent; and *P*=0.02 RRR=0.49 for multiple-changing). This suggests that carriage group based on individual *spa* types may be classifying some individuals with persistent carriage and evolution in the nose as multiple-persistent carriers which could be reducing the association between single-persistent carriage and older age.

4.4.5 Investigation of carriage duration

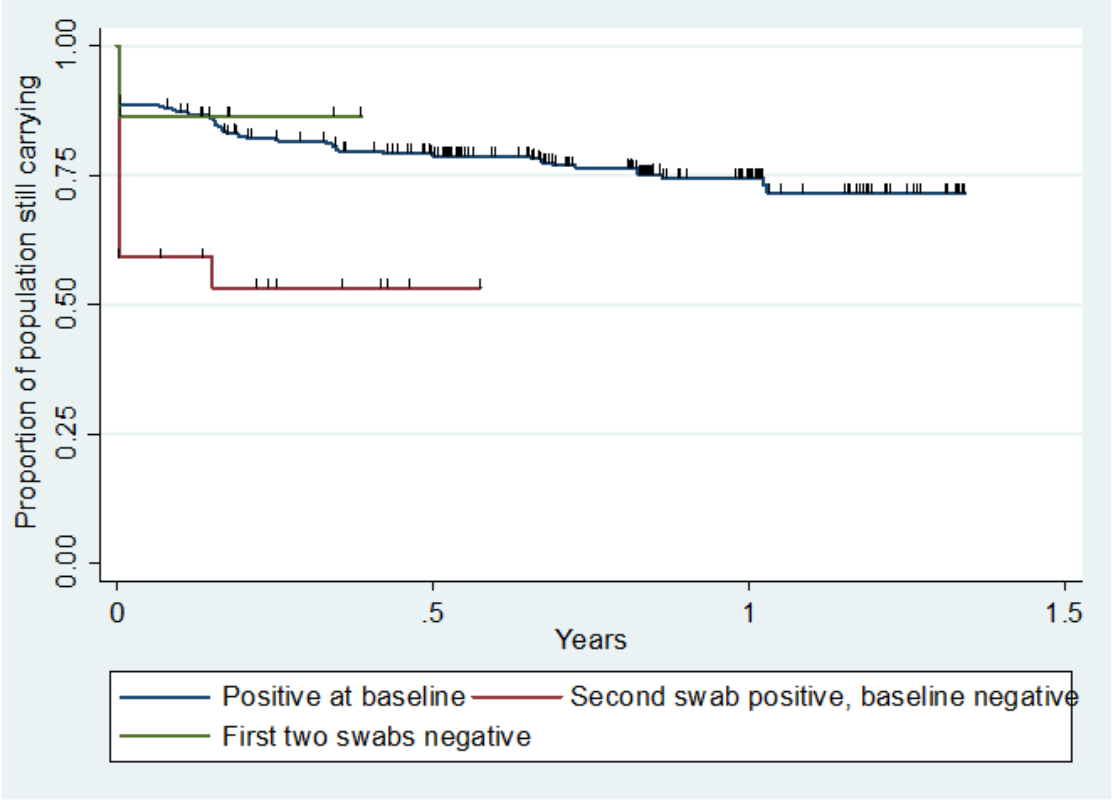
4.4.5.1 Individual carriage duration

As an alternative to investigating longitudinal carriage by grouping types of carriage into discrete categories, carriage duration was also calculated using positive and negative follow-up data combined. Kaplan Meier survival curves were used to estimate rates of decay of carriage, firstly at the level of the individual, where two consecutive negative swabs were considered loss of carriage. Individuals with a negative final swab and positive second to last swab had unknown carriage duration, since it was not known whether the final swab was a false negative, therefore they were censored at the second to last swab and, thus, not included in the loss of carriage group. Carriage duration was calculated separately for three groups: (i) individuals positive at baseline (i.e. all those in positive follow-up); (ii) individuals with baseline negative and second swab positive (as a single negative swab might be a false negative); and (iii) individuals with at least first two swabs negative.

Figure 4.9 is a Kaplan Meier plot showing the proportion of individuals from the above three groups carrying *S. aureus* as time since baseline increases. A drop in the line indicates loss of carriage when a positive swab is followed by two swabs negative for *S. aureus*. A dash on the line indicates that an individual has no follow-up after this time, so it is unknown whether they continued to carry *S. aureus*. The immediate drop at one day represents loss of carriage after one swab. Overall, *S. aureus* carriage appears to last for over one year in the majority of people. Kaplan Meier plots are normally used to calculate median durations from the point at which there is failure (in this case loss of carriage) in 50% of subjects. However, 50% of participants did not lose carriage during the study period so it was not possible to calculate median carriage duration. Instead 25th percentile carriage

length was calculated, which was 315 days for individuals positive at baseline. In individuals with first swab negative and second swab positive the 25th percentile carriage duration was a single swab only, and only 14% of individuals with first two swabs negative lost carriage, so it was not possible to calculate the 25th percentile survival time for them.

Figure 4.9: Carriage duration at the level of the individual



Individuals with a positive baseline swab carried for significantly longer with a lower rate of loss (log rank test *P*=0.003) compared to individuals with baseline or first two swabs negative (Figure 4.9). However, Cox regression demonstrated that the significant difference in carriage duration was due to the difference in loss of carriage after only one swab (Table 4.16). Here, individuals with baseline negative and second swab positive lost carriage after one swab significantly more often compared to individuals positive at baseline, whereas the rate of loss for subsequent swabs did not differ between groups.

Table 4.16: Cox regression of carriage duration at the level of the individual between groups, separated by those with a single positive swab and those with multiple positive swabs

	Hazard Ratio	LCI	UCI	*P* value
Carriage of a single positive swab:				
Second swab positive baseline negative versus baseline positive	3.07	1.37	6.85	0.006
First two swabs negative versus baseline positive	1.39	0.43	4.48	0.59
Carriage of at least two subsequent positive swabs:				
Second swab positive baseline negative versus baseline positive	1.09	0.15	8.00	0.93
First two swabs negative versus baseline positive	No individuals			

4.4.5.2 Carriage duration at the strain level

Carriage duration was considered at the strain level to detect *S. aureus* in the nose changing over time, and therefore re-classify participants who were persistent carriers at the individual level, but in whom the strains carried were changing. Two consecutive swabs negative for a specific *spa* type were considered loss of carriage. Figure 4.10 is a Kaplan Meier plot similar to figure 4.9; a drop represents a loss of *spa* type, where a *spa* type on one swab is followed by two swabs negative for that *spa* type. As for investigation of carriage duration at the level of the individual, it was not possible to calculate median carriage duration for all groups. Only in individuals with first swab negative and second swab positive did more than half of participants lose carriage, with median carriage duration of 28 days. However, 25th percentile carriage duration could be calculated for all three groups and was 83 days for those positive at baseline, and a single swab for those with first, and first and second swab negative. The lower quartile carriage duration for individuals positive at baseline at the level of the *spa* is therefore less than half the time of the lower quartile duration at the level of the individual.

Comparison between the three groups (positive at baseline; second swab positive and baseline negative; and first two swabs negative) again showed that they had significantly different carriage durations (log rank test between three groups $P<0.001$) (Figure 4.10). However, the significant difference in carriage duration was again due to the difference in loss of carriage after one swab rather than the difference in carriage duration of more than one swab as demonstrated by Cox regression (Table 4.17).

Table 4.17: Cox regression of carriage duration at the level of the strain between groups separated by those with a single instance of a *spa* type and those with multiple subsequent identical *spa* types

	Hazard Ratio	LCI	UCI	*P* value
Carriage of a single *spa* type:				
Second swab positive baseline negative versus baseline positive	2.81	1.73	4.57	<0.001
First two swabs negative versus baseline positive	1.98	1.26	3.09	0.003
Carriage of at least two subsequent identical *spa* types:				
Second swab positive baseline negative versus baseline positive	1.98	0.94	4.15	0.07
First two swabs negative versus baseline positive	0.29	0.04	2.11	0.22

Figure 4.10: Carriage Duration at the level of the strain

Three individuals with first two swabs negative with maximum carriage duration for that group and carriage loss were censored at 295 days.

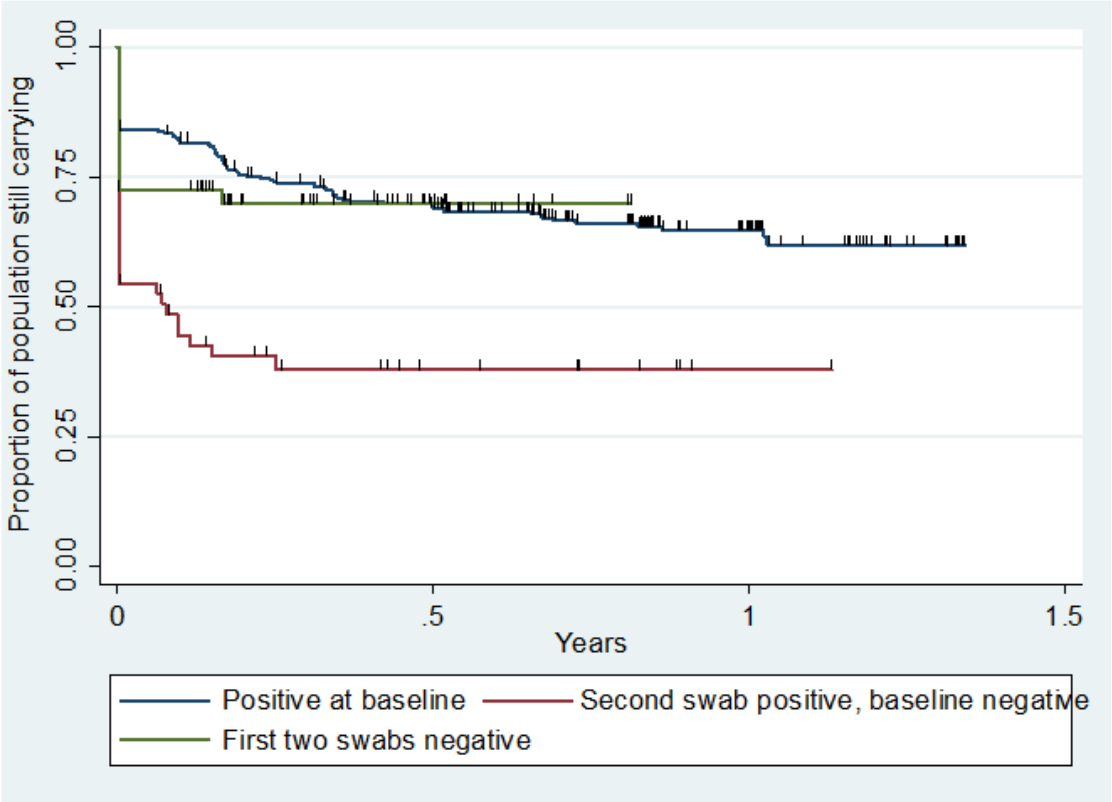


Table 4.18: Rate of loss per year at the level of the individual and strain

	Level of the individual		**Level of the strain**	
Group	N	Rate *S. aureus* loss per year (number carriage lost/year)	N	Rate strain loss per year (number strain lost/year)
Positive baseline	360	0.4	374	0.6
Positive second swab, negative baseline	22	2.8	57	2.8
Negative first two swabs	22	1.5	117	2.6
All above	404	0.4	548	0.8

Carriage duration, measured by both the lower quartile time to loss of carriage and the rate of carriage loss (Table 4.18) was greater at the level of the individual than the strain. This is explained by persistent carriers who change *spa* type over time, and are categorised differently at the level of the individual and the strain. However, the extent of the difference between carriage duration calculated at the level of the individual and the level of the strain may be a slight overestimate.

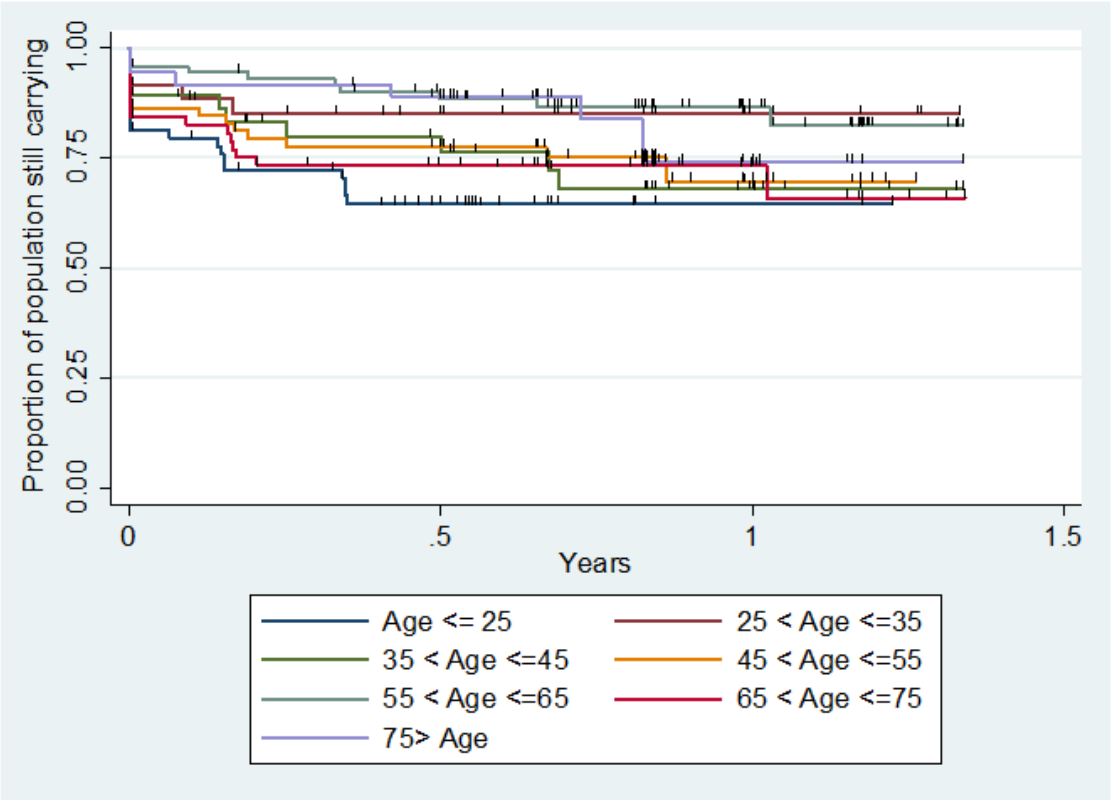
Twenty-eight percent of individuals classified as multiple-persistent carriers carried *spa* types within the same CC which may have arisen through evolution in the nose, therefore a change in *spa* type for these individuals may not represent loss of carriage.

For individuals positive at baseline, loss of carriage was low with more than half of individuals carrying for over one year. However, individuals negative at baseline lost carriage more rapidly with double the rate of loss per year, again suggesting that negative follow-up comprises individuals who are more likely to be transient carriers. Additionally carriage of a single positive swab was observed more often at the level of the *spa* than the level of the individual, which could indicate temporary colonisation with a second strain not seen during the rest of carriage.

4.4.6 Host environmental factors and carriage duration

Since both male sex and age were associated with *S. aureus* carriage status in this study, their relationship with carriage duration was tested. Kaplan Meier curves were calculated at the level of the individual for those in positive follow-up only (since individuals in negative follow-up were a small group with different carriage behaviour) and separated by male and female, or divided into 10 year age groups. Males and females did not carry for significantly different durations (P=0.17), however, the different age groups did carry for significantly different durations (P=0.04); individuals under 25 years old had shorter carriage durations (Figure 4.11). This finding is in keeping with younger individuals in negative follow-up being more likely to have *S. aureus* isolated from at least one follow-up swab, suggesting that younger people are more likely to be transient carriers.

Figure 4.11: Carriage duration at the level of the individual, separated by age group



4.4.7 Longitudinal carriage and bacterial genetics

To investigate whether *S. aureus* lineage had an effect on persistence or duration of carriage, *spa* types were grouped into CCs using BURP clustering and isolates in each *spa* CC found in more than 10 individuals were tested to determine whether they were distributed evenly between carriage groups using Fishers' Exact Test. This identified two *spa* CCs: CC012 and CC032, that were carried significantly more as a single-persistent type (*P*=0.04 and *P*=0.02 respectively), and one *spa* CC: CC024 carried significantly more as a multiple-persistent type and less as a single-persistent type (*P*=0.02) (Table 4.19).

Table 4.19: Distribution of isolates from each CC into carriage groups and *P* value representing whether the distribution is even across groups

CC	Total	Single-persistent	Single-intermittent	Multiple-persistent	Multiple-changing	*P* value
CC002	23	9	8	4	2	0.32
CC012	126	56	22	33	15	0.04
CC015	29	8	6	10	5	0.47
CC024	24	3	4	11	6	0.02
CC032	31	17	4	10	0	0.02
CC078	14	2	3	8	1	0.17
CC084	53	15	8	18	12	0.46
CC089	13	5	0	5	3	0.23
CC127	29	15	4	5	5	0.32
CC136	11	7	1	0	3	0.07
CC148	14	8	3	1	2	0.25
CC659	12	3	0	3	6	0.09

To support the finding that isolates in CC22 and CC30 are more likely to be carried singularly and persistently MFP AIC regression was performed investigating the effect of being in CC012 and CC032 (combined to increase numbers) or CC024 compared to any other CC, on carriage group assignment. Regression revealed that isolates in CC012 and CC032 were less likely to be carried as a single-intermittent, multiple-persistent or multiple-changing type compared to carriage as a single-persistent type (*P*=0.045 RRR=0.57; *P*=0.04 RRR=0.61; and *P*=0.001 RRR=0.32 respectively) (Table 4.20). As opposed to this, isolates in CC024 were more likely to be carried as a multiple-persistent or multiple-changing type compared to carriage as a single-persistent type (*P*=0.02 RRR=4.86; and *P*=0.03 RRR=4.68 respectively) (Table 4.20). Kaplan Meier analyses of carriage duration of isolates positive at baseline also suggested a trend where isolates in CC012 and CC032 were carried for longer and CC024 were carried for shorter durations, however neither of the differences were significant (*P*=0.06 and *P*=0.6 respectively).

Table 4.20: Multivariate model for the effect of being in CC012 or CC032 and CC024 compared to other CCs on carriage type

	RRR	LCI	UCI	*P*
CC0122 or CC032				
Single-intermittent versus single-persistent	0.57	0.33	0.99	0.045
Multiple-persistent versus single-intermittent	0.61	0.38	0.98	0.04
Multiple-changing versus single intermittent	0.32	0.001	0.17	0.61
CC024				
Single-intermittent versus single-persistent	2.68	0.58	12.26	0.20
Multiple-persistent versus single-intermittent	4.86	1.32	17.80	0.02
Multiple-changing versus single intermittent	4.68	1.14	19.24	0.03

CC012 and CC032 are the *spa* CCs associated with the two most common UK healthcare associated STs: ST36 and ST22 respectively (seen in Chapter Three). That these healthcare-associated *spa* types are more likely to be carried as a single-persistent type may help to explain their success as they may be able to out-compete other *S. aureus* strains. The difference was not attributable to age, since individuals with strains in *spa* CC012 and CC032 did not have different ages from individuals with *spa* types in other CCs. *spa* CC024 includes *spa* types related to ST8, associated with USA300 CA-MRSA in the USA. Further investigation of the isolates in *spa* CC024 that were carried as a multiple-persistent type, revealed 9/10 included *spa* types in more than one CC, suggesting that *S. aureus* in *spa* CC024 is more likely to be transiently carried.

4.5 Discussion

Results from the study were consistent with the hypothesis that *S. aureus* carriage in the community is correlated with environmental host risk factors and bacterial genetic factors. When investigating carriage cross-sectionally four host environmental factors were identified as likely predictors of *S. aureus* carriage: male sex, current employment; participation in contact sport; and over 30 days since most recent out-patient appointment. Furthermore, investigation of longitudinal carriage revealed a relationship between *S. aureus* longitudinal carriage and age, with younger participants more likely to carry multiple *spa* types intermittently; to carry for shorter durations; and to have *S. aureus* isolated from subsequent swabs if negative at baseline. Finally, when investigating bacterial genetics and longitudinal carriage, epidemic hospital strains in CC012 and CC032 were carried significantly more often as a single-persistent type.

The study also provided further investigation of the hypothesis that there is CA-MRSA in Oxfordshire, including investigation of individuals who did not have samples sent into hospital for routine testing.

Nine of 1123 (0.8%) individuals were found to carry MRSA, of which two fit the CDC definition for CA-MRSA, however these both had healthcare associated *S. aureus* strains and were isolated from individuals with other healthcare risk factors. Therefore MRSA identified in the community is largely healthcare associated, and could represent hospital strains that have gone feral in the community, as seen in Chapter Three.

4.5.1 Cross-sectional carriage rates

As part of investigating cross-sectional carriage, the study was able to elucidate *S. aureus* carriage rates in Oxfordshire. Similar to *S. aureus* prevalence in other studies, almost one third of participants in this study (32%) carried *S. aureus* (Kluytmans, van Belkum et al. 1997; Van den Bergh, Yzerman et al. 1999). This suggests a stable carrier phenotype as the prevalence of *S. aureus* in the community remains the same as observed 50 years ago (Williams 1963). MRSA carriage rates were much lower than suggested by the majority of previous studies (Tiemersma, Bronzwaer et al. 2004). However, MRSA rates are often calculated in at risk populations with increased MRSA risk factors, so MRSA carriage is likely to be lower in healthy individuals. Comparison of MRSA carriage rates in this study with a study of the general population in the USA revealed an identical MRSA colonisation rate of 0.8% (Kuehnert, Kruszon-Moran et al. 2006).

4.5.2 Longitudinal carriage

The majority of individuals in positive and negative follow-up continued to carry *S. aureus* until their last swab. This made it impossible to calculate median carriage duration directly, however the available data is consistent with typical carriage durations of at least one year and possibly much longer (Van den Bergh, Yzerman et al. 1999). Categorisation into carriage groups revealed that 20% of participants enrolled into the study were persistent carriers (of either single or multiple *spa* types), a value consistent with other findings (Van Belkum, Riewarts Eriksen et al. 1997; Van den Bergh, Yzerman et al. 1999; van den Akker, Nouwen et al. 2006). However, investigation of the *spa* type of each strain, including co-colonisation of multiple strains in the nose meant that longitudinal carriage could be classified in much more detail than previous studies.

Investigation of individuals negative at baseline found 20% had *S. aureus* growth from at least one subsequent swab. Few studies have been performed on *S. aureus* acquisition in the healthy population, however, this value is similar to results found in a small study looking at *S. aureus*

acquisition in healthy individuals where 30/177 (17%) acquired *S. aureus* (Sakwinska, Blanc et al. 2010).

4.5.3 Host environmental factors and carriage

After consideration of detailed information correlated with *S. aureus* cross-sectional carriage versus non-carriage, the most significant predictor was male sex, a well known risk factor (Wertheim, Vos et al. 2004). The three other likely predictors of *S. aureus* carriage (current employment; participation in contact sport; and over 30 days since most recent out-patient appointment) had not been found in previous studies investigating *S. aureus* carriage. It is interesting that both current employment and contact sport were significant in a multivariate model as these risk factors were both associated with younger age (both $P<0.005$), which was significant in univariate analysis but not included in the multivariate model. This suggests that the univariate relationship between age and *S. aureus* carriage was due to the behaviour of younger people, since it is probable that the additional close contact with peers both of these activities provide increased the likelihood of *S. aureus* acquisition.

Despite collection of detailed epidemiological and healthcare data, only age was associated with longitudinal carriage. The lack of correlation with any other factor was supported by a previous study of *S. aureus* carriage where age, gender, patient contact and hospitalisation were not correlated with retaining, losing or acquiring *S. aureus* (Sakwinska, Blanc et al. 2010). In this study GP records will be re-visited after two years of follow-up, allowing collection of details about events that occurred since recruitment. It is likely that these may correlate with carriage behaviour, thus revealing additional factors associated with longitudinal carriage.

The association between *S. aureus* carriage and age showed young people are more likely to carry transiently. They are more likely to carry multiple *spa* types intermittently, to carry for shorter durations, and to have *S. aureus* isolated from subsequent swabs if negative at baseline. It was expected that a difference in carriage would be seen in children which is why the study only recruited participants 16 years and older, the maximum age for paediatrics in the UK. However even in young adults, an association remained. The effect of age on carriage persistence and duration may have been even stronger if recruitment of younger participants was more successful, and younger age groups were better represented.

Many studies of *S. aureus* carriage have identified an age effect whereby younger individuals have higher carriage rates (Noble, Valkenburg et al. 1967; Armstrong-Esther 1976; Bischoff, Wallis et al. 2004; Graham, Lin et al. 2006; Kuehnert, Kruszon-Moran et al. 2006; Munckhof, Nimmo et al. 2009; Qu, Cui et al. 2010), some of which particularly investigated young adults. For example, it has been suggested that carriage is particularly high in individuals aged 10 to 20 (as referenced by (Williams 1963)), and undergraduates have been demonstrated to have higher carriage rates than children under 16 or adults over 19 years (McFarlan 1938). However, other studies conflict with these findings (Kuehnert, Kruszon-Moran et al. 2006; Halablab, Hijazi et al. 2010).

Very few studies have investigated age and type of carriage. Intermittent carriers have been demonstrated to be more likely to carry multiple types of *S. aureus* however this relationship has not been associated with age (Van Belkum, Riewarts Eriksen et al. 1997). A comparison of carriage in infants under one year old and adults found that infants were significantly more likely to carry more than one type of *S. aureus* (Peacock, Justice et al. 2003) defined by MLST and PFGE. Also, a study of individuals aged 55 years and above found younger individuals to be associated with persistent carriage compared to non-carriers, however, the carriage phenotype was defined from only two nasal swabs (Emonts, Uitterlinden et al. 2008). Interestingly, two studies have failed to find an association between age and intermittent or persistent carriage (Noble, Valkenburg et al. 1967; Eriksen, Espersen et al. 1995) and conflicting results have reported a change from persistent to transient carrier status in 10 to 19 year olds (Armstrong-Esther 1976). However, no previous studies have looked at the type of *S. aureus*, only presence or absence, so although carriage persisted at the individual level the strain may change more often in children.

Transient carriage in younger participants could be explained by social behaviour, as they may be more likely to come into close contact with other individuals, and therefore exchange *S. aureus* with them. This is in agreement with participation in contact sport and current employment being risk factors for *S. aureus* carriage at baseline, and previous research has also demonstrated that social events are a risk factor for *S. aureus* carriage (Qu, Cui et al. 2010). This is particularly the case for the cohort of new university students investigated in negative follow-up; however, the relationship between acquisition and age remained when students were not included in the analysis.

4.5.4 Bacterial genetics and carriage

Three *spa* CCs were distributed significantly differently between carriage types, two of which (CC012 and CC032) are associated with epidemic hospital strains ST22 and ST36 and were found to be carried significantly more often as a single-persistent type. The association between hospital type strains and single-persistent carriage has not been found before, since no studies have followed carriage, including *spa* typing every sample, over an extended period. Many isolates in *spa* CC012 and CC032 remained persistently carried to the 18 month end point for analysis, and it is possible that stronger associations will be found once the study is completed after two years, and perhaps even longer.

4.5.5 Limitations and compromises

Time constraints mean that this initial analysis covers only the first 18 months since the beginning of recruitment in December 2008. Therefore, the maximum time participants were followed was 18 months and a median of only seven swabs were returned. At the end of the study participants will have returned a maximum of 14 swabs over a two year period, and information will be collected again from GP records to correlate healthcare behaviour with carriage over time. This may reveal further relationships between host environmental factors or bacterial genetics and *S. aureus* longitudinal carriage not seen with this subset of the data. Furthermore, it is unfortunate that recruitment into negative follow-up began later than positive follow-up so less data has been collected from the negative individuals. As the study continues, the small proportion of individuals in negative follow-up who currently have all but baseline positive swab positive may lose carriage. Additionally, increasing duration of time in negative follow-up may mean more participants acquire *S. aureus*. In fact, if the suggestion by Van den Bergh, that all non-carriers will acquire *S. aureus* when the environmental conditions are appropriate (Van den Bergh, Yzerman et al. 1999) is correct, then if followed for long enough, all individuals in negative follow-up may eventually acquire *S. aureus*.

Due to low numbers of young (particularly male) individuals attending GP practices it was not possible to recruit participants with an age and sex distribution representative of the general population, even after recruiting university students to increase numbers in the lowest age group. In particular, participants in negative follow-up were an unrepresentative group, since they consisted of 128/210 (61%) students who were found to be more likely to grow *S. aureus* from their nasal swabs if they started the study negative. Results from this group are therefore less applicable to the general

population. Ten percent of participants in negative follow-up from Bampton had *S. aureus* growth from at least one swab, which is lower than found in previous research (Sakwinska, Blanc et al. 2010) however, this previous study was performed on hospital employees, so levels may be higher than the rate in the general population. Additionally, once individuals in the negative follow-up arm of this study have been followed for longer, a larger proportion may have *S. aureus* identified from at least one swab.

Despite detailed collection of risk factors, this study did not investigate all possible risk factors for *S. aureus* carriage. The study was designed to investigate *S. aureus* from nasal colonisation, surface swabs and sterile sites; as well as progression to *S. aureus* disease; and to compare risk factors between these groups. Risk factors investigated were therefore chosen from studies looking at MRSA. This meant that some factors previously associated with *S. aureus* carriage were not investigated, namely: smoking, obesity (Herwaldt, Cullen et al. 2004), washing of the nose (Halablab, Hijazi et al. 2010) and pets (Huijsdens, van Dijke et al. 2006). Smoking, obesity and use of a nasal spray (washing of the nose) are recorded in GP records and will be collected in a second review of GP records after two years. Pets are not recorded reliably in GP records, therefore details will be collected using the form sent to participants with the nasal swab pack, which they will be asked to complete and return with their swabs.

Measures other than demographic or healthcare risk factors may also be associated with *S. aureus* carriage. Firstly, other bacteria growing in the nose have been shown to affect *S. aureus* growth (Iwase, Uehara et al. 2010) and therefore may affect whether people carry *S. aureus*. It would be possible to use nasal swabs to investigate this, but the study protocol aimed to quickly identify *S. aureus* as part of routine laboratory culture using enrichment broth and selective chromogenic agar specifically for *S. aureus* identification. Identification of other bacteria would make laboratory work much more time consuming as conditions could not be optimised for growth of *S. aureus*.

Similarly, previous studies have demonstrated that the amount of *S. aureus* in the nose affects whether an individual is an intermittent or persistent carrier (van Belkum, Verkaik et al. 2009; Sakwinska, Blanc et al. 2010). However, measuring *S. aureus* colony forming units would require standardisation of nasal swabbing, which was not possible with self swabbing. Additionally, identification of *S. aureus* using PCR may have revealed extra individuals with very low levels of

S. aureus colonisation, not identified using culture based methods. However culture based identification allowed each sample to be stored as a glycerol stock, and prevented false positive results from individuals with dead *S. aureus* cells in the nose.

Finally, host genetics may affect *S. aureus* carriage. Various studies have shown different host factors to be correlated with nasal carriage (van den Akker, Nouwen et al. 2006; Emonts, Uitterlinden et al. 2008; Sivaraman, Venkataraman et al. 2009). Investigation of this would require sequencing of human DNA, which is beyond the scope of this project, however an ethics application is being prepared by members of the Modernising Medical Microbiology consortium to carry out such a study.

4.5.6 Impact of results

Correlation of three new host environmental factors with cross-sectional *S. aureus* carriage (current employment; participation in contact sport; and over 30 days since most recent out-patient appointment), provides three new ways to identify individuals who are more likely to carry *S. aureus*. Such individuals could be targeted for altered or more rapid treatment, since they are known to be predisposed to carriage. This could be of particular importance on admission to hospital, especially if the individual has a skin breach, which could allow their carried *S. aureus* to enter the blood and cause infection.

Investigation of longitudinal carriage revealed that the majority of individuals carry *S. aureus* for at least one year. Therefore, individuals who acquire *S. aureus* in hospital are likely to continue to carry and transmit this strain for quite some time after discharge, and on re-admission, people with previous hospital exposure over a year ago are still at risk of *S. aureus* carriage. Additionally, the correlation of younger people with more transient carriage means that their carriage status is more likely to have changed from a previous result, so perhaps indicates they require more frequent testing, for example if in hospital. It may also mean that younger adults are more likely to acquire *S. aureus* on admission to hospital, thus may require extra precautions whilst in hospital. However risk factors present within hospital differ from healthy carriage in the community so whether the relationship between age and carriage remains in the hospital environment would first need to be confirmed.

The association of strains in CC012 and CC032 with single persistent carriage could explain the success of healthcare associated types, which may be able to out-compete other types in the nose, hence being more likely to be identified singly, and once in the nose may be able to survive for longer.

Displacement of *S. aureus* in the nose by another strain has not been demonstrated *in vivo*, however experiments have demonstrated that certain *S. aureus* strains are able to outgrow others *in vitro* (Laurent, Lelievre et al. 2001). Further experiments may be required to determine the causal factor of this association.

Finally, the study was able to confirm the low levels of CA-MRSA in Oxfordshire. Only 2/1123 individuals had MRSA which fit the CDC definition for CA-MRSA, and even these were nosocomial type strains. This suggests that, CA-MRSA in the UK is uncommon and not currently causing the threat seen in the USA (Groom, Wolsey et al. 2001; Hussain, Boyle-Vavra et al. 2001; Moran, Krishnadasan et al. 2006; Chambers and Deleo 2009).

4.5.7 Future work

This chapter represents only a small part of a larger study that was designed to investigate *S. aureus* invasive and superficial disease as well as carriage. Therefore once participants from other parts of the study have been recruited it will be possible to investigate the hypothesis that: *S. aureus* superficial and invasive disease is correlated with environmental host risk factors and bacterial genetic factors. Once the set of disease isolates is collected comparisons will also be possible between the carriage and disease isolates, in order to investigate the hypothesis that: environmental host factors and bacterial genetics differ between *S. aureus* carried in the community and causing disease.

As well as the hypotheses generated from the other parts of this study, additional hypotheses arose from the results of this study. It is likely that host genetics and competition from other co-colonising bacteria are also correlated with *S. aureus* carriage. Investigation of further hypotheses such as: *S. aureus* carriage in the community is associated with host genetics; and co-colonisation of the nose with other bacteria, could help to further explain the differences in carriage between individuals.

Although the study found associations between age and bacterial genetics with carriage over time, the causal mechanism for each of these effects was not identified. Investigation of the immune response of *S. aureus* carriers may reveal differences with age that explain why younger people are more likely to carry transiently. Alternatively, the effect of carriage and age may be explained by social behaviour which could be investigated by asking participants more detailed questions about their social interaction and investigating their friends and partners. Such investigations would answer

hypotheses including: the change in carriage behaviour with age is caused by changes in host immunity or social behaviour.

Finally, whole-genome sequencing of the *S. aureus* strains could reveal genes related to the single-persistent nature of strains in CC012 and CC032, and answer hypotheses such as: *S. aureus* in CC012 and CC032 carry genes predisposing them to persistent carriage. Various samples from healthy carriers in the study are currently being whole-genome sequenced and will provide more detail on the population structure of *S. aureus* carried in Oxfordshire. Comparison of genes present in samples with different longitudinal carriage would allow an association type study to answer part of the above hypothesis. Furthermore, once individuals with *S. aureus* disease have also been recruited it will be possible to use whole-genome sequencing to investigate hypotheses such as: *S. aureus* causing disease differs genetically from *S. aureus* that is carried.

5 Diversity in CC22 and CC30 over space and time

5.1 Introduction

'Next-generation sequencing' is a term that describes recent techniques developed to generate DNA sequence data using methods other than traditional Sanger sequencing (Sanger, Nicklen et al. 1977) (section 1.9.1). Next-generation sequencing methods parallelise sequencing by allowing millions of DNA sequences to be produced at once, thus sequences are produced much faster and also at a lower expense. There are several currently available next-generation platforms, including Illumina, 454 and SoLiD (Shendure and Ji 2008) (section 1.9.1), of which the Illumina platform has been most widely adopted. Furthermore, several new methods are becoming available which will soon make it possible to sequence whole pathogen genomes within a few hours in an individual lab (Travers, Chin et al. 2010; Chin, Sorenson et al. 2011).

As next-generation technologies continue to develop, population-scale genomic analyses are becoming more common. Several studies of the evolution of bacterial species or groups within species have already been published including studies on *S. aureus* (Harris, Feil et al. 2010), *Streptococcus pneumonia* (Croucher, Harris et al. 2011), *Bacillus subtilis* (Srivatsan, Han et al. 2008), and *Mycobacterium tuberculosis* (Gardy, Johnston et al. 2011). Additionally, a further study using third-generation sequencing to investigate *Vibrio cholera* (Chin, Sorenson et al. 2011) has now been published. This chapter will use next-generation sequencing to investigate the diversity and evolutionary history of *S. aureus* in two CCs (CC22 and CC30) common in UK hospitals.

5.2 Hypotheses

The study was designed to investigate the following hypotheses:

Sequence diversity exists within populations of a single *S. aureus* MLST

CC22 and CC30 have similar levels of diversity

CC22 and CC30 sequence diversity is spatially and temporally structured

5.2.1 Importance of the hypotheses

The study was designed to investigate *S. aureus* in CC22 and CC30 since, in the UK, the bulk of serious *S. aureus* disease is caused by samples from these two clonal lineages. Therefore all isolates sequenced were from CC22 and CC30, which include epidemic hospital strains ST22 and ST36

respectively. Together ST22 and ST36 caused 96% of *S. aureus* bloodstream infections in 2000 (Johnson, Aucken et al. 2001), thus the majority of nosocomial MRSA is indistinguishable using conventional molecular typing techniques. It is therefore important to understand more about evolution and history of these STs which may provide insight into their success within UK hospitals.

Furthermore, similar STs are commonly grouped into CCs. Whole-genome sequencing will enable investigation of the true extent of within CC diversity, beyond the seven MLST alleles, allowing identification of evolutionary relationships within CCs. Additionally, investigation of two CCs will allow comparison of diversity between them, in order to reveal whether CCs contain similar diversity, which is important if findings are to be generalised to other *S. aureus* CCs. Within CC diversity will also be partitioned into spatial-temporal groups as isolates in each CC were taken from two geographic locations and two time periods separated by about ten years, during which samples within these CCs were the predominant cause of MRSA bacteraemia (Ellington, Hope et al. 2010). Detailed investigation of these CCs will enable understanding of the spread of epidemic *S. aureus* strains between hospitals in relation to the diversity created in one place over time, allowing investigation of the rise of epidemics over time and their transmission between hospitals.

Results generated from investigating the above hypotheses could also be used to investigate differences between carriage and disease isolates. Some isolates used in this study were initially collected to compare carriage with community and hospital-associated disease. This is important to investigate, since despite many studies investigating putative virulence factors (Table 1.1) (Day, Moore et al. 2001; Day, Moore et al. 2002; Feil, Cooper et al. 2003; Lindsay, Moore et al. 2006), currently only TSST shows a clear association (Tristan, Ferry et al. 2007) with *S. aureus* disease and even that is not a determining factor. It is possible, however, that the resolution provided by whole-genome sequencing may finally reveal some factors explaining why some *S. aureus* causes disease. Due to the complex nature of *S. aureus* disease, it is unlikely that a single genetic factor will be found causing all strains to have a distinct phenotype, however even identification of a factor associated with increased disease risk in some strains would be a major breakthrough in the prevention of *S. aureus* disease. Such associations between genes and phenotype may allow a targeted approach to preventing and removing strains from individuals that are more likely to be harmful or could enable rapid diagnostic tests to identify disease causing strains and lead to appropriate treatment.

5.2.2 Novelty of the hypotheses

Next-generation sequencing is a relatively new technology providing answers to many previously unanswerable questions. Since designing this study, within ST diversity has been demonstrated for 63 isolates of ST239 (Harris, Feil et al. 2010). However, this study is the first to investigate diversity of an entire CC, within which two sets of isolates from individual STs were also available, both larger than any *S. aureus* sample sets sequenced before. Findings from this study will therefore be able to confirm results from the ST239 study, whilst allowing further generalisation to other STs. This study is the first to use whole-genome sequencing to investigate UK epidemic CCs, so will provide a novel insight into two highly relevant lineages, the majority of which could not be previously be distinguished between using molecular typing. Furthermore, whole-genome sequencing has never been used to compare *S. aureus* isolates from specific geographic and temporal groups, thus results of this study represent the first comparison of *S. aureus* diversity created over space and time.

5.3 Methods

In order to investigate the hypotheses, collections were assembled of isolates from CC22 (n=101) and CC30 (n=199) spanning two locations within England: Oxford and Brighton, and two time periods separated by approximately 10 years. Collections included both MSSA and MRSA as well as carriage and disease samples. The collections were whole-genome sequenced using the Illumina sequencing platform and analysed using high-throughput bioinformatics methods.

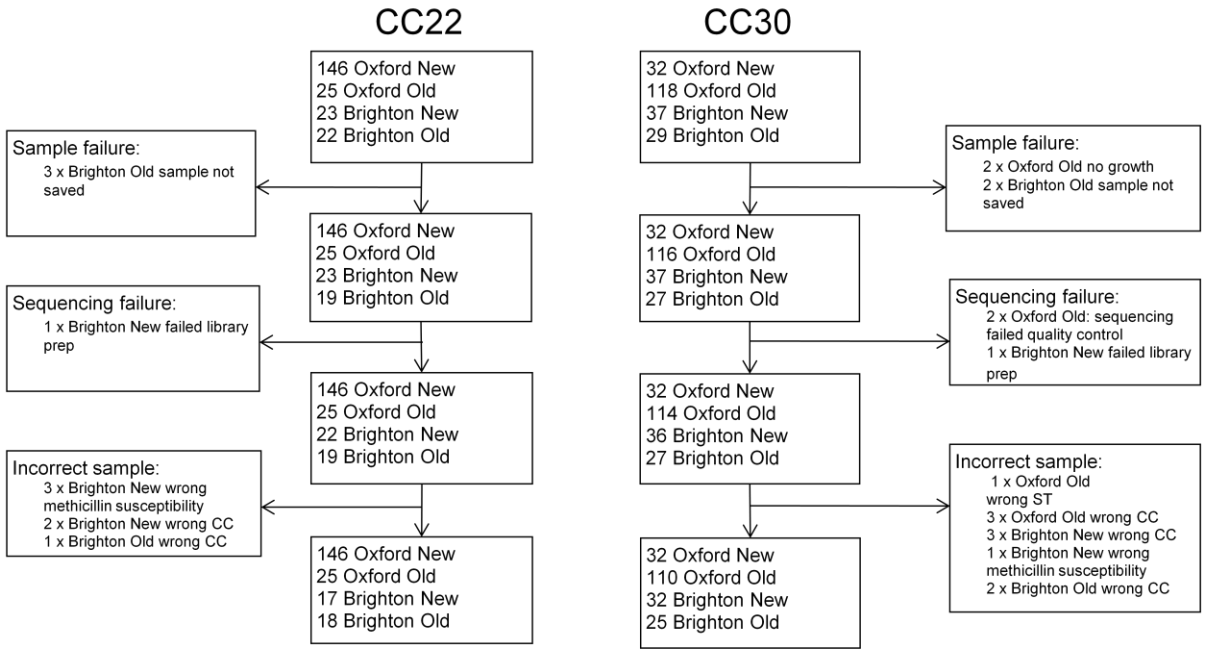
5.3.1 Sample selection

Isolates were taken retrospectively from previous *S. aureus* collections and divided into four groups: Oxford Old, Oxford New, Brighton Old and Brighton New. “Oxford Old” isolates were a collection of invasive disease and healthy carriage samples taken between 1997 and 1998 (Peacock, Moore et al. 2002) and “Oxford New” isolates from the sample sets used in Chapter Three of this thesis, collected between 2003 and 2007. “Brighton Old” and “Brighton New” isolates were both from collections of *S. aureus* bacteraemia between 1999 and 2001; and 2006 and 2007 respectively (Price, Baker et al. 2010).

From the available collections all possible isolates were sequenced, apart from CC22 Oxford New MRSA where 146 isolates were available and 40 were randomly chosen for sequencing, since 40 was larger than any other group of MRSA isolates sequenced previously. Figure 5.1 shows a flow chart of

sample processing. Isolates were removed from each group for the following reasons: five were not saved (all Brighton Old); two did not grow from culture (both Oxford Old); two failed library prep and were not repeated (both Brighton New); and two failed sequencing (both Oxford Old) and were not repeated.

Figure 5.1: Sample selection flow chart for each CC



Once sequenced, further isolates were removed as the characteristics of their resulting DNA sequence were not as expected. Four isolates were actually MRSA when expected to be MSSA (all Brighton New), initially identified when the DNA sequence output did not include the *mecA* gene, and verified by standard sensitivity testing of the isolates according to the Methods. STs were obtained from the Illumina output for each isolate (section 5.3.3.8), from these one Oxford Old isolate had the wrong ST within the correct CC and three Oxford Old isolates had STs within the wrong CC. As Oxford Old isolates were sequenced first, these were re-sequenced, and subsequent Illumina results confirmed that the initial isolates were incorrectly labelled, thus were excluded. Additionally, eight further isolates (five Brighton New and three Brighton Old) also had sequence from the wrong CC and were excluded.

Forty isolates were successfully sequenced more than once (one five times, 15 four times, three three times; 21 twice). From each set of replicates, only the isolate with the greatest amount of sequence data (defined as number of reads) was selected. One-hundred and one isolates in CC22 and 199

isolates in CC30 were included in the final analysis. Table 5.1 gives the distribution of isolates in each clonal complex into spatial-temporal groups.

Table 5.1: Number of isolates (and number which were MRSA) sequenced in each CC from each time period and geographic location

	CC22		CC30	
	Oxford N (MRSA)	**Brighton**	**Oxford**	**Brighton**
New	41 (40)	17 (16)	32 (25)	32 (24)
Old	25 (4)	18 (16)	110 (23)	25 (14)

The study was conducted under the limitations of retrospective sampling in order to contrast, as effectively as possible, time and space in the evolution and diversity of the two dominant CCs of UK *S. aureus*. The result was that compromises had to be made in the balance of other variables. Therefore, samples from each spatial-temporal group were not distributed evenly with respect to disease presentation and place of acquisition (Table 5.2). In particular, isolates from Brighton came from two studies focused on hospital-acquired disease, so none were carriage or community samples.

Table 5.2: Distribution of isolates in (a) CC22 and (b) CC30 from each spatial-temporal group by disease presentation and place of acquisition

(a) CC22	Carriage	Disease	Hospital	Community	Total
Oxford Old	9 (22%)	32 (78%)	34 (83%)	7 (17%)	41
Oxford New	14 (56%)	11 (44)%	19 (76%)	6 (25%)	25
Brighton Old	0 (0%)	18 (100%)	18 (100%)	0 (0%)	18
Brighton New	0 (0%)	17 (100%)	17 (100%)	0 (0%)	17

(b) CC30	Carriage	Disease	Hospital	Community	Total
Oxford Old	53 (48%)	57 (52%)	69 (63%)	41 (37%)	110
Oxford New	8 (25%)	24 (75%)	8 (25%)	24 (75%)	32
Brighton Old	0 (0%)	25 (100%)	25 (100%)	0 (0%)	25
Brighton New	0 (0%)	32 (100%)	32 (100%)	0 (0%)	32

5.3.2 Ilumina sequencing

Sequencing was undertaken by the Genomics Group at the Wellcome Trust Centre for Human Genetics (WTCHG), Oxford (http://www.well.ox.ac.uk) using the Illumina TruSeq chemistry on Illumina GAIIx and HiSeq2000 machines. The project spanned developments in the chemistry, software and hardware used for sequencing, presenting additional challenges to analysis. Output from Illumina is in multiple overlapping DNA fragments or “reads” which vary in length. Sequences of isolates included in

the final CC22 and CC30 datasets had read lengths between 51-100 bases long. Output was provided in fastq format.

Reads were assembled by mapping to a CC-specific reference, which offers several advantages over *de novo* assembly (section 1.9.2). *De novo* assembly produces output in many fragments that are difficult to align and compare, whereas assembly by mapping to a reference produces one complete genome that is easily aligned to all others assembled to that reference. Additionally, if the reference is well annotated it allows easy identification of genes in the isolate sequenced. Also, use of a CC-specific reference increased the quality of the mapping, since mapping quality increases with the similarity of the sample to the reference. In this study isolates in CC22 were mapped to EMRSA15: an ST22 isolate from Imperial college London, sequenced by the Sanger; and isolates in CC30 were mapped to MRSA252: an ST36 isolate from Oxford, which was part of the Oxford Old dataset, sequenced by the Sanger.

Reference based assembly does not, however, identify anything present in the sample sequenced but not the reference, which may include MGEs with important phenotypic consequences. The extent of information lost is reduced by mapping samples to a more similar reference genome, as was the case in this study, when isolates were mapped to a CC-specific reference. However, it remains likely that some additional regions of the isolate sequenced will not be identified. To further identify genes present in the isolates sequenced but not in the CC-specific reference, reads were also mapped to an experimental "pan-genome" constructed from all 23 *S. aureus* reference genomes and 49 *S. aureus* plasmids available from NCBI (as accessed on 01/02/2011). Despite the care taken to include all known elements from *S. aureus* the formal possibility remains that elements present in the samples sequenced may not be present in the pan-genome, these would only be identifiable using *de novo* assembly.

5.3.3 Bioinformatics methods

Generation of complete sequence data from Illumina sequence output; and further analysis of the sequence data in order to answer the hypotheses, was a multi-step process requiring use of many different software packages and bespoke scripts. The software and scripts used are outlined below.

5.3.3.1 Maq

Maq 0.7.1 (Li, Ruan et al. 2008) (http://maq.sourceforge.net/) was used to align reads to a reference genome, when sequence output had read lengths of 36-51 bases. Maq parameters were all set at default, except the minimum outer distance for two paired reads was 1000 bases; and the fraction of heterozygotes allowed among sites was 0.0.

5.3.3.2 Stampy

Stampy v1.0.11(Lunter and Goodson 2010) (http://www.well.ox.ac.uk/project-stampy) was used to align reads to a reference genome when sequence output was over 100 bases, since Maq does not support output of 100 base reads. Stampy was used with default parameters for all options, except that Burrows-Wheeler Aligner (BWA)-pre-alignment was omitted and the expected substitution rate was set at 0.01, not 0.001. To ensure consistency for analysis, all data initially mapped using Maq was remapped using Stampy.

5.3.3.3 Samtools

After mapping to a reference using Stampy base calls of each isolate with respect to the reference genome were inferred using the MPILEUP program of the SAMTOOLS package v0.1.12-10 (r896) (http://samtools.sourceforge.net/).

5.3.3.4 Arlequin

Once whole genomes had been assembled for each of the isolates, Arlequin (Excoffier, Laval et al. 2005) (http://cmpg.unibe.ch/software/arlequin3/) was used to quantify the diversity within CCs. Arelequin provided the population average pairwise difference within (π_x) and between (π_{xy}) groups of isolates (section 1.8.5) and was used for calculation of these values within STs and CCs as well as between spatial-temporal groups.

5.3.3.5 ClonalFrame

To account for the effect of mutation and recombination on tree topology, the program ClonalFrame (Didelot and Falush 2007) (http://www.xavierdidelot.xtreemhost.com/clonalframe.htm) was used to draw phylogenetic trees of isolates from each CC. ClonalFrame uses a Markov chain Monte Carlo (MCMC) algorithm to estimate the genome-wide clonal relationships between sequences. It identifies recombination events from segments of individual genomes where there are more variant positions than expected given the distribution of mutation events over the entire genome. ClonalFrame jointly

estimates both the history of the samples and the parameters of the mutation and recombination processes.

ClonalFrame was designed for use on datasets created using MLST, which are much smaller than whole-genome sequence data, therefore two modifications had to be made for use with the CC22 and CC30 datasets. Modifications were: reduction in the number of iterations used to estimate the parameters (the tree, the location of recombination events on the branches of the tree, and the mutation and recombination rates), which on post-hoc examination did not affect parameter estimates; and reduction of the time taken for each iteration, achieved in two ways. Firstly, it was assumed that uncalled sites, represented by Ns, in positions where only one actual base had been called were not polymorphic. Secondly, the number of "reference sites" which ClonalFrame defines as polymorphic sites or sites at positions divisible by 50 was reduced, because reference sites are calculated more slowly than the sites between them. Here reference sites were reduced to be only sites divisible by 100 rather than 50. With these modifications, ClonalFrame took 10 and 20 days to analyse CC22 and CC30 respectively. Dr Xavier Didelot made the modifications to ClonalFrame for this dataset, and produced the resulting parameters from the ClonalFrame output.

5.3.3.6 Mauve

Mauve (Darling, Mau et al. 2010) (http://gel.ahabs.wisc.edu/mauve/) is used to produce multiple genome alignments and was required for construction of a pan-genome. Mauve produces multiple outputs including a multi-fasta file showing the alignment of sequences entered and their rearrangements. It also produces a "backbone" file listing all the contigs from each of the sequences entered and which contigs were present in each sequence. The backbone file was required for construction of the pan-genome.

5.3.3.7 Xbase and Basys

Two freely available programs were compared for genome annotation of the pan-genome: Xbase (http://www.xbase.ac.uk/annotation/) and BASys (http://basys.ca/basys/cgi/submit.pl). BASys (Van Domselaar, Stothard et al. 2005) was chosen to annotate the pan-genome because it identified more coding sequences in each genome, and annotated more coding sequences with a gene and product name.

5.3.3.8 Bespoke scripts

The majority of bespoke scripts were written by myself (unless stated) using Python (http://www.python.org/), including Biopython (/http://biopython.org/wiki/Biopython) for analysis of sequence data. Bespoke scripts used included determination of STs from whole-genome sequence data; generation of pairwise difference matrices; calculation of the d_N/d_S ratio; and assembly of the pan-genome. Details of each are outlined below.

To determine STs, pileup files were made from the Stampy output, which gave details of every base mapped to every position in the reference. From the pileup file a Python script was used to identify the most common base mapped to each position, which was deemed the consensus call. A further Python script was then used to slice the consensus sequence from the genome at the position of each of the MLST genes. The sequence of each gene from each isolate was submitted to the MLST database as described in the Methods.

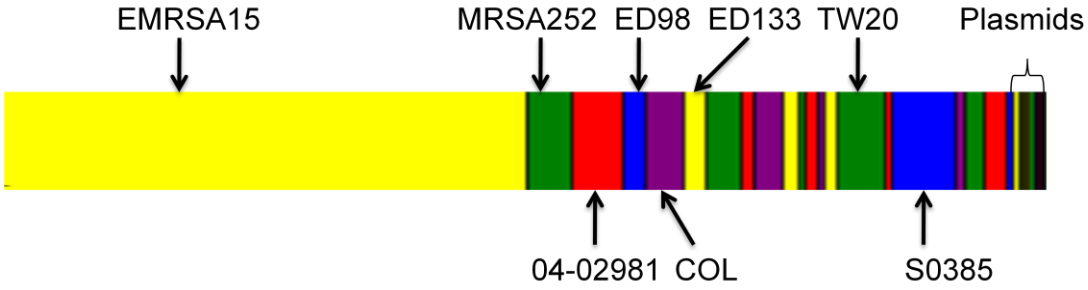
Pairwise difference matrices were generated to compare and depict the diversity within each CC and its predominant ST. Each matrix was made using a Python script, written by Dr Madeleine Cule, which calculated the number of variant positions between each pair of isolates included in the matrix (not including uncalled positions). To depict the variation within the matrices, heat maps were drawn using a modified version of an R script, also written by Dr Madeleine Cule.

The d_N/d_S ratio was calculated to indicate whether a region of DNA was under directional or purifying selection (section 1.8.2). In brief, d_N is the number of non-synonymous point mutations and d_S the number of synonymous mutations. A ratio of greater than one is suggestive of positive selection, whereas a ratio of less than one indicates purifying selection, although because every gene is likely to contain a mixture of sites with different evolutionary constraints and d_N/d_S represents the average across these sites, it is just a guide. In this study the d_N/d_S ratio was calculated by generating a consensus sequence for each CC including all positions that varied from the reference. A Python script written by Dr Tanya Golubchik was used to determine whether variants in coding regions were synonymous or non-synonymous, and these were summed using a further Python script. The final ratio was divided by three, since almost exactly three times as many non-synonymous point mutations as synonymous mutations are possible, assuming equal codon usage.

Construction of the pan-genome began by aligning all 23 *S. aureus* reference genomes and 49 *S. aureus* plasmids available from NCBI (as accessed on 01/02/2011) using Mauve and taking output from the backbone file (section 5.3.3.6). It was then possible to combine one copy of each contig to make a pan-genome, by assembling all the contigs from the first genome in order, followed by appending all the contigs in the second genome not present in the first in order and so on. To increase the quality of the mapping, all contigs from the first genome were replaced with the entire genome sequence from the CC-specific reference, after which all other contigs were appended. There were therefore two CC-specific pan-genomes, for CC22 and CC30 starting with EMRSA15 and MRSA252 respectively, which were 5577564 and 5520389 bases long, approximately twice as long as a single *S. aureus* genome (Figure 5.2). Reads were mapped to the pan-genome using Stampy (section 5.3.3.2) and the same approach taken for the mapping to CC-specific references. Each pan-genome was annotated using Basys (section 5.3.3.7) to enable information about genes and gene products to be determined for positions where isolates aligned to the pan-genome.

Figure 5.2: CC22 pan-genome

Each coloured block represents a different reference genome or plasmid. Labels indicate the names of some of the reference genomes the pan-genome consists of.



Analysis of the pan-genome was performed only on the region beyond the end of the initial CC-specific reference genome sequence, since this represented a complementary approach to the initial mapping to a CC-specific reference. The non-reference pan-genomes included 3658 and 3558 genes in the CC22 and CC30 specific pan-genomes respectively (according to the Basys annotation), which were not present in the CC-specific reference but present in one or more other reference genomes. These non-reference regions consisted entirely of elements from the accessory genome, although not exclusively, since the CC-specific references also contained some accessory genes. Any mention of mapping to the pan-genome in this study will only consider the second half, past the specific CC reference, i.e. the non-reference region.

Mapping to the pan-genome was analysed using a contig based approach, chosen because each isolate only contained a small minority of the accessory genes represented on the non-reference pan-genome, and no accessory genes were present in all isolates. The majority of mapping therefore consisted of Ns (99.5% and 99.4% of mapping to the CC22 and CC30 non-reference pan-genomes respectively). Contigs were defined as any region where more than 200 bases were mapped, with no gaps greater than five bases within the contig after filtering. A contig length of 200 was chosen, since this, combined with the rigorous filtering parameters (section 5.4.2.4), was sufficient to ensure contigs were not mapping artefacts, and produced no contigs in the non-reference pan-genome when the CC30 reference was mapped to the CC30-specific pan-genome.

5.4 Results

5.4.1 Sequence assembly

CC-specific reference genomes were used so the vast majority of reads (mean 97% for both CC22 and CC30) were mapped to their specific reference. Tables 5.3 and 5.4 outline summary statistics regarding the success of sequencing and mapping of each of the isolates in the dataset. Quality is negative-log-transformed (Phred-like) error rates and was similar for both CCs.

Table 5.3: CC22 summary statistics for sequencing, mapping and filtering

	Median (IQR)	**Maximum**	**Minimum**
Reads	3485808 (2792444, 5126848)	22549380	2155474
Reads mapped	3405515 (2720975, 4939637)	2006806	22023970
% reads mapped	97.9% (97.5%, 98.0%)	99.5%	89.0%
Read depth	72 (31, 96)	253	21
Quality	256 (147, 285)	307	109

Table 5.4: CC30 summary statistics for sequencing, mapping and filtering

	Median (IQR)	**Maximum**	**Minimum**
Reads	3566122 (2813469, 5259100)	24972884	1318070
Reads mapped	3472626 (2752583, 5121428)	24824735	1255054
% reads mapped	97.2% (96.2%, 98.2%)	99.6%	86.2%
Read depth	70 (31, 95)	251	14
Quality	254 (176, 287)	323	97

Mapping to the pan-genome resulted in a higher percentage of reads mapped than were to the CC-specific references. A mean of 98.6% for CC22 and 98.4% for CC30 reads were mapped, which was a mean of 66412 and 55452 more reads mapped per isolate respectively. This represented the extra regions in the genomes of the isolates not present in the CC-specific references.

5.4.2 Filtering

Raw output in the Maq and Stampy consensus files had a high false-positive rate for calls deemed different from the reference, so it was necessary to filter the results. In order to develop and evaluate filters, two sets of identical samples (sa45-29 and sa21-18) that had been repeatedly sequenced to a high depth, yielding many reads aligned to each position in the reference, were divided into 12 pseudo-samples. Both sets of pseudo-samples were ST36, so were mapped to the ST36 reference (MRSA252) using Maq. Filters were optimised to maximise the number of true variants that passed the filter, whilst minimising false positives. Here “true variants” were considered to be positions where all 12 pseudo-samples differed from the MRSA252 reference before filtering and “false positives” were positions where some but not all of the 12 pseudo-samples differed from the reference.

5.4.2.1 Quality based filtering

Quality-based filtering was initially designed using output from Maq. It considered various measures from the SNP file (containing quality information about each position in the reference) and pileup files (containing information about each base aligned to each position in the reference) produced by Maq. Information considered included: read depth (the number of reads aligned to a certain position in the reference); Maq quality score; whether the base was likely to be in a repetitive region in the reference (indicated by a high self-self Basic Local Alignment Search Tool (BLAST) similarity) or sample; quality of surrounding bases; proportion of calls the same as the consensus call; and proportion of calls the same as the reference call.

Various combinations and values of each of these parameters were investigated. The final filter was based on the following six parameters at each position in the reference: (i) read depth greater than two; (ii) Maq quality score greater than four; (iii) Maq quality score greater than 1.2 * read depth; (iv) quality of the surrounding bases greater than 19; (v) proportion of calls the same as the consensus greater than 75%; and (vi) when read depth is less than ten, all calls must be the same as the consensus. Calculation of pairwise differences between fasta files produced for each of the 12

pseudo-samples showed a mean of 27 base differences between each pair in sa45-29 and eight base differences between each pair in sa21-18. Results were higher for sa21-18, since sa45-29 contained a repetitive region that was harder to map.

5.4.2.2 Probabilistic Goodness of Fit filter

A second filter using a probabilistic goodness of fit method was developed by Dr Daniel Wilson as an alternative filtering method to the quality based filter. It was again designed to maximise the number of true variants from the pseudo-samples passing the filter, whilst minimising other variant calls. The probabilistic filter calculated the proportion of reads with the same call as the reference genome for each position in the reference. Then, for each position in the reference it calculated the likelihood that the position was different from the reference given the distribution of the number of calls the same as the reference in the rest of the genome. The goodness of fit filter also called positions as “N” if there was not enough information to make an accurate call.

Compared to the quality based filter, the probabilistic goodness of fit filter called fewer variants between each pair of 12 pseudo-samples, with a mean of 0.2 and 0.05 variants per pairwise comparison for sa45-29 and sa21-18 respectively. However, this was at the cost of having some unknown calls (Ns), since 178412 and 584028 positions or 6% and 20% of the genomes were called as N in sa45-29 and sa21-18 respectively.

5.4.2.3 UK Clinical Research Collaboration (UKCRC) Modernising Medical Microbiology (MMM) data pipeline filter

Finally, incorporating knowledge from the quality-based and goodness of fit filters, the MMM data pipeline filter was developed. The MMM data sequence pipeline was developed by the Modernising Medical Microbiology Consortium to streamline the process of receiving raw sequence output from Illumina and producing completed DNA sequence for each sample sequenced. The MMM data pipeline filtered the raw sequence output using the following parameters: (i) minimum read depth of five with at least one read in each of the forward and reverse direction; (ii) minimum depth not less than the 2.5% of the read depth distribution for the sample; (iii) maximum depth not greater than the highest 2.5% of the distribution for the sample; (iv) minimum root-mean-square read mapping quality of 30; (v) all samples with bases aligned to a particular position in the reference must have a minimum mean base quality of 10; (iv) minimum of 75% of reads supporting the consensus call; (v) read not in

a repetitive region in the reference indicated by a high self-self BLAST similarity; (vi) sites not called in positions where there were two indels or three variants within 12 bases; and; (vii) calls must have sufficient support for a single base call under the diploid model assumed by SAMTOOLS.

Sensitivity and specificity of the MMM pipeline was calculated by comparing sequencing results from the same sample that had been sequenced multiple times, rather than the two test sets of 12 identical pseudo-samples. Forty isolates for this study were sequenced more than once. Pairwise comparisons between each set of duplicate isolates revealed zero called differences between any of the pairs, demonstrating the very high specificity of the MMM pipeline filter. The cost for the high specificity of the MMM pipeline filter was a median 259485 or 9% Ns in each of the duplicate isolates. The MMM pipeline filter was used for the rest of this study.

After filtering each isolate in CC22 and CC30 using the MMM pipeline filter, sequences had a median of 202637 uncallable positions (Ns) in CC22 and 283752 in CC30, which represented 7.2% and 9.8% of the reference genome lengths respectively. CC30 had significantly more N's (rank-sum $P<0.005$), probably because isolates in CC30 were more divergent from the reference (mean 1534 differences) than CC22 (mean 672 differences), suggesting that CC30 was slightly harder to map.

5.4.2.4 Modifications of the MMM filter for the pan-genome

Extra confidence was required to ensure that reads mapping to the pan-genome beyond the CC-specific reference truly represented additional genes. Therefore the minimum depth value for the MMM filter in the part of the pan-genome beyond the CC-specific reference was increased to be greater than half the depth of the mapping to the CC-specific reference at the start of the pan-genome. Additionally, the MMM filter was altered so it did not remove repetitive regions in the part of the pan-genome past the CC-specific reference, since they were likely to represent the edge of Mauve contigs.

5.4.3 Diversity within a single MLST

In order to investigate diversity with a single ST, it was first necessary to confirm the MLST of each isolate sequenced. Whole-genome sequencing offers a new way to reliably perform MLST without need for PCR and sequencing reactions, allowing comparison of samples sequenced using next-generation technologies to others where only molecular typing has been performed. It was possible to reliably call MLST using unfiltered Illumina output, since MLST uses sequences of housekeeping

genes, which are easy to sequence using Illumina because they are non-repetitive and have low levels of variation.

Each CC was dominated by one UK epidemic ST: 93% of CC22 was isolates from ST22 and 42% of CC30 was isolates from ST36 (Table 5.5). These two STs were further examined to investigate within ST diversity. Notably, ST2046 and ST2047 both represent new STs found in this study, identified solely from the Illumina output. ST2046 had a single base difference from ST22 in the gene *yqiL* and ST2047 a single base difference from ST36 in gene *pta*.

Table 5.5: Frequency of each ST in the CC22 and CC30 datasets

CC22		**CC30**	
ST	**Frequency (%)**	**ST**	**Frequency (%)**
21	1 (1%)	2	1 (0.5%)
22	94 (93%)	24	1 (0.5%)
23	1 (1%)	30	70 (35%)
44	1 (1%)	31	1 (0.5%)
60	1 (1%)	32	1 (0.5%)
134	1 (1%)	33	1 (0.5%)
957	1 (1%)	34	14 (7%)
2046	1 (1%)	36	84 (42%)
		37	1 (0.5%)
		38	1 (0.5%)
		39	20 (10%)
		41	1 (0.5%)
		43	1 (0.5%)
		57	1 (0.5%)
		2047	1 (0.5%)

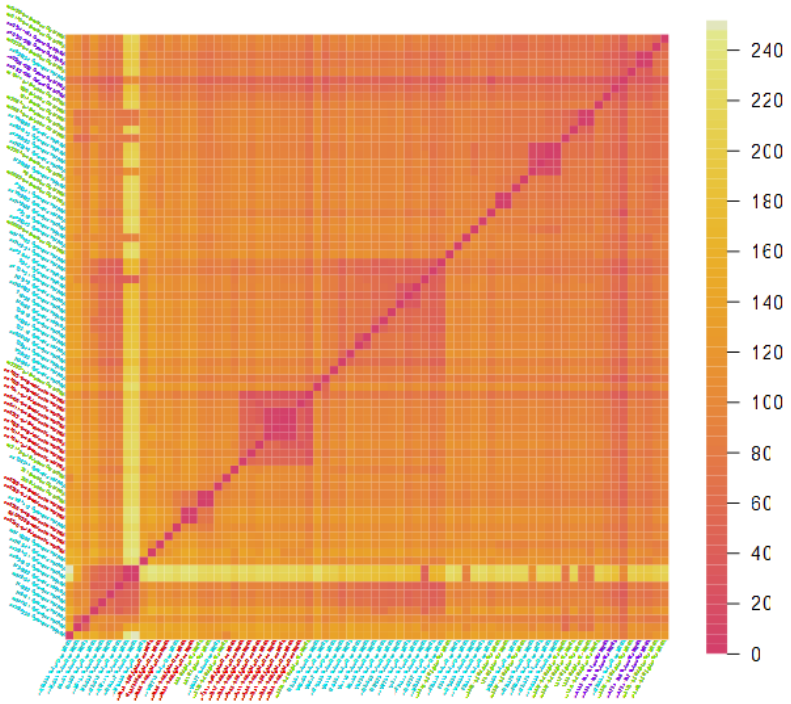
To determine the extra variability revealed by whole-genome sequencing compared with MLST and standard laboratory susceptibility testing variation between MRSA isolates in the two predominant STs in each CC, ST22 and ST36, was investigated. In total, between the 94 isolates in ST22 there were 2318 variant positions and there were 2508 variant positions between the 84 isolates in ST36. Pairwise difference matrices were calculated for MRSA isolates in each ST, from which heat maps were drawn to depict the extra diversity that was identified within isolates indistinguishable by MLST. In the heat map the number of variant positions between each isolate was graded on a colour scale, with red showing the most similar isolates and yellow the most divergent (Figure 5.3).

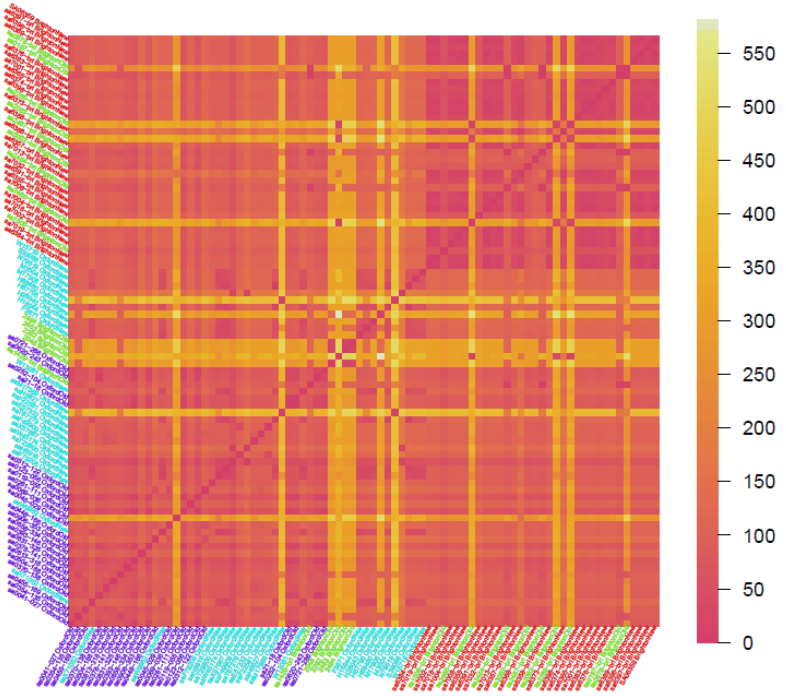
The heat maps in figure 5.3 show substantial diversity within STs with a maximum number of variant positions between two isolates of 581 for ST36 and 252 for ST22. The mean number of variant

positions between isolates (π_x) was 97 in ST22 and 148 in ST36, suggesting that isolates indistinguishable by MLST are likely to differ by at least 100 bases. Despite this, even after sequencing the entire genome there was one pair of identical isolates found in ST36, both from patients in the Brighton New group, isolated two months apart, with identical antimicrobial susceptibility patterns.

Figure 5.3: Heat map of variant positions between every isolate in ST22 (left) and ST36 (right)

Each square represents the number of variant positions between two isolates on a colour scale as represented to the right of each heat map (note the greater scale in ST36). Isolates are ordered according to position in a UPGMA tree. Blue legends are Oxford New, Purple Oxford Old, Red Brighton New and Green Brighton Old.





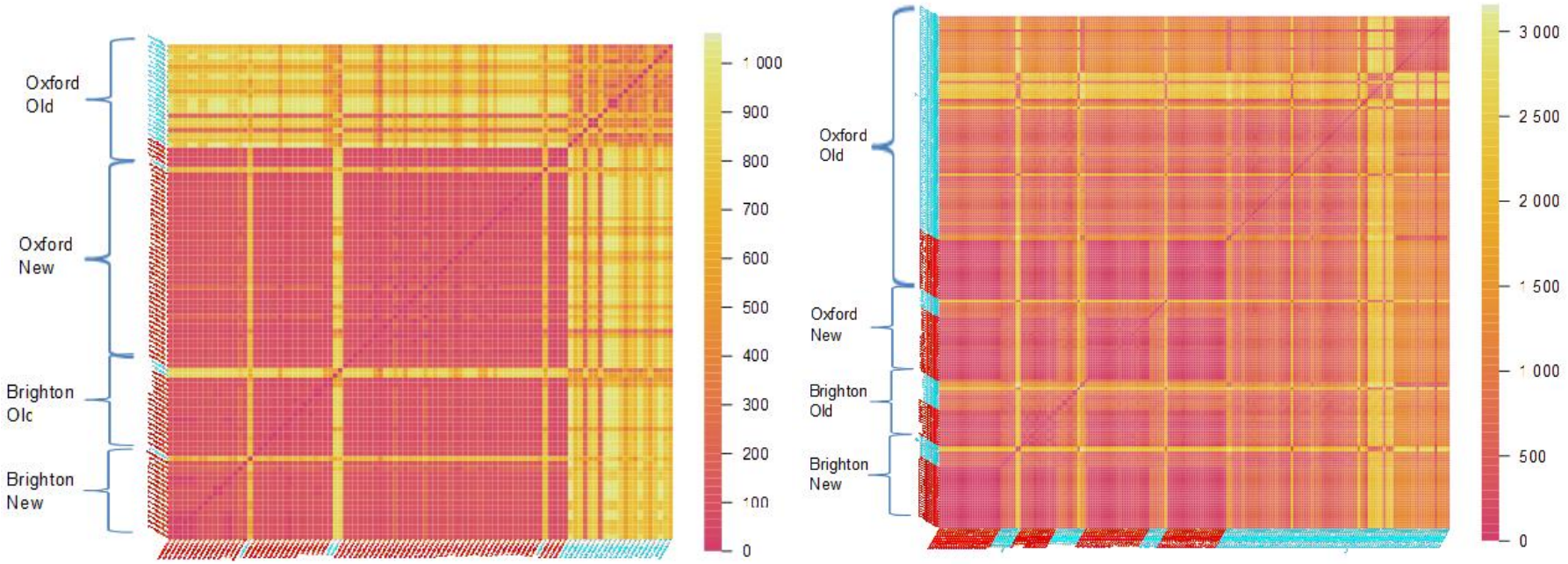
5.4.4 Diversity within CCs

In total between the 101 isolates in CC22 there were 7318 variant positions and between the 199 isolates in CC30 there were 23957 variant positions. Matrices of pairwise differences for each CC revealed that the maximum number of variant positions between two isolates was 1063 for CC22 and 3161 for CC30. CC30 also had a greater population average pairwise difference (π_x) with 853 nucleotide differences between isolates in CC30 compared to 367 for CC22, suggesting that CC30 has increased diversity compared to CC22. To depict the variation within each CC two further heat maps were drawn (Figure 5.4). Additionally, a further pair of identical isolates was also identified in CC30, both of which were Brighton New MSSA isolates in ST30, isolated one day apart with identical susceptibility patterns. All other isolates within each CC could be differentiated by at least one variant position.

Comparisons of π_x for MRSA and MSSA isolates in each CC showed that MRSA was much less diverse than MSSA. MSSA isolates in CC22 and CC30 had a π_x of 469 and 1073 pairwise differences respectively, as opposed to only 101 and 158 pairwise differences for MRSA isolates in each CC. The increased similarity of MRSA isolates is highlighted in the heat maps in figure 5.4.

Figure 5.4: Heat map of variant positions between every isolate in CC22 (left) and CC30 (right)

Each square represents the number of variant positions between two isolates on a colour scale as represented to the right of each heat map (note the greater scale in CC30). Isolates were ordered according to spatial-temporal group and within this by meticillin sensitivity and ST. Red legends are MRSA and blue legends MSSA.



5.4.4.1 Phylogenetic trees

Phylogenetic trees were drawn using the output from ClonalFrame (section 5.3.3.5), representing the relationships between the genomes at the level of substitutional diversity (Didelot and Falush 2007). Figures 5.5 and 5.6 show ClonalFrame trees for CC22 and CC30 respectively.

Figure 5.5: ClonalFrame phylogenetic tree for CC22

Terminal branches are coloured according to whether the sample was meticillin-sensitive or resistant. Scale bar indicates the average number of mutation - recombination events that occur during the interval of time represented by the scale.

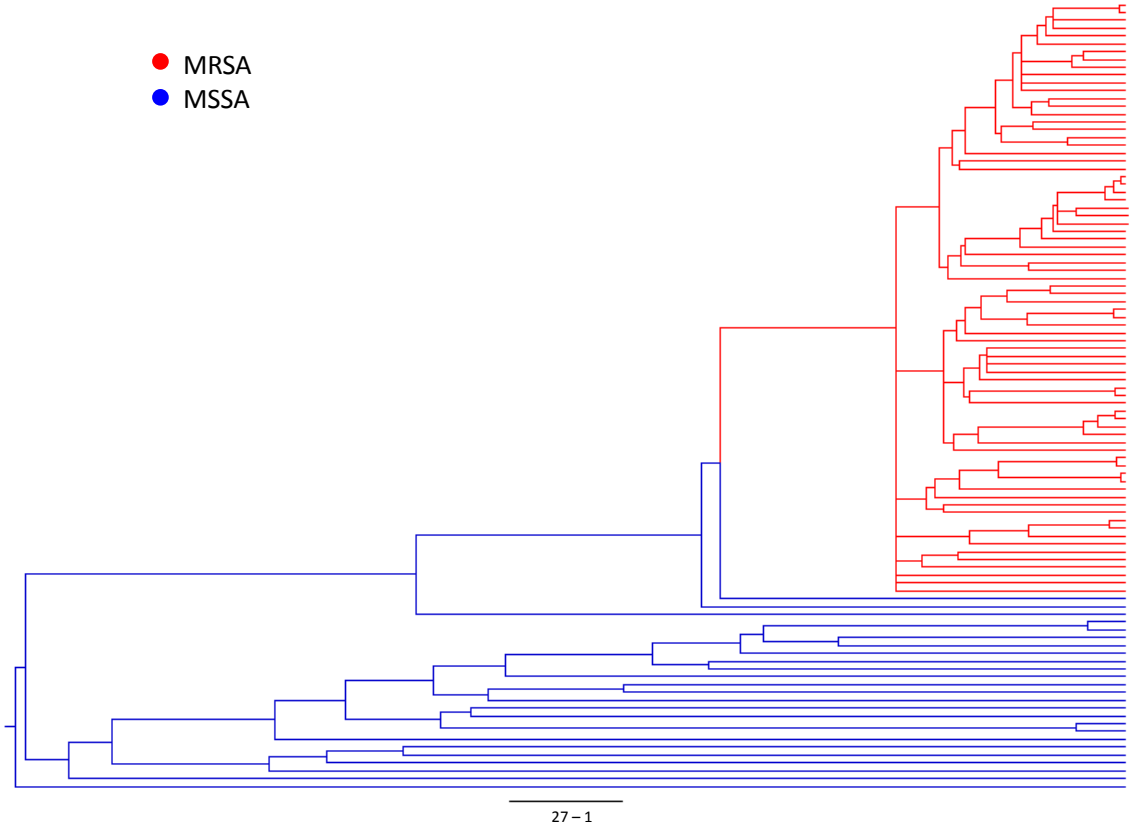
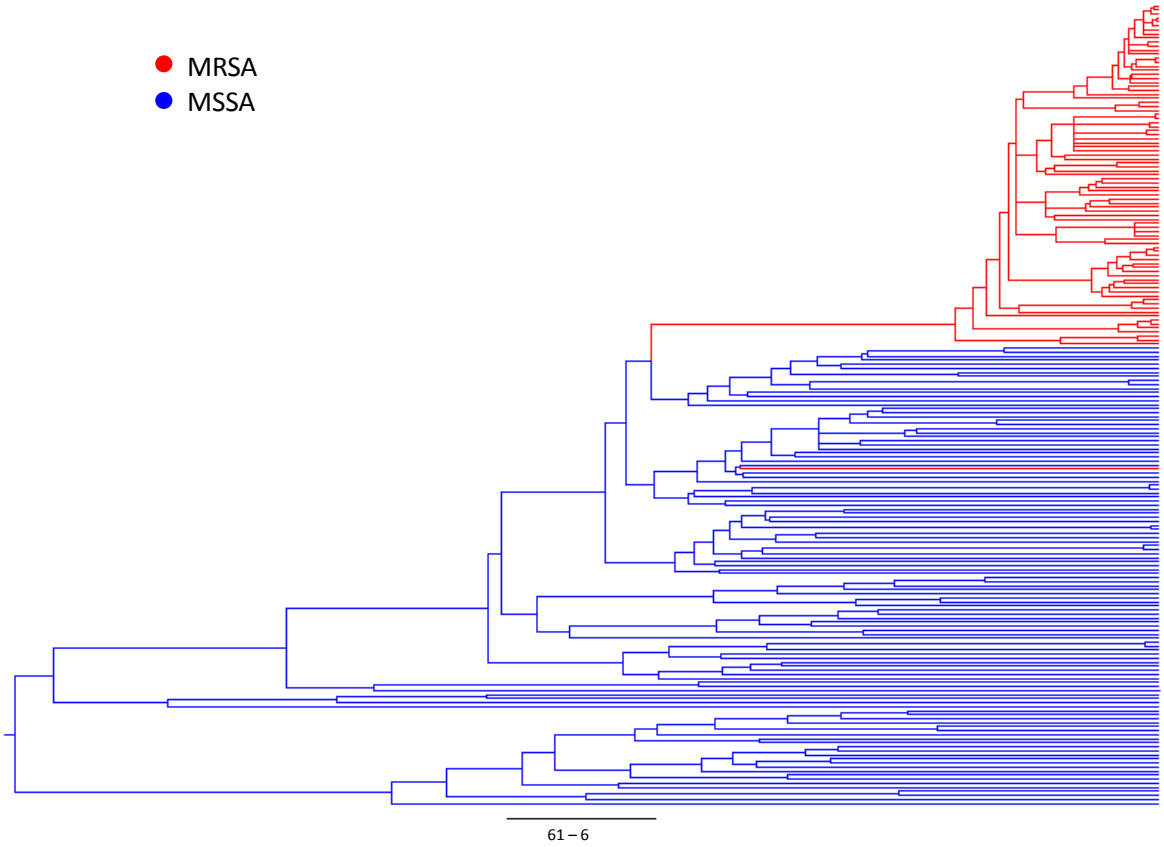


Figure 5.6: ClonalFrame phylogenetic tree for CC30

Terminal branches are coloured according to whether the sample was meticillin-sensitive or resistant. Scale bar indicates the average number of mutation - recombination events that occur during the interval of time represented by the scale.



The ClonalFrame trees emphasise the greater diversity of CC30 compared to CC22. The scale on each tree indicates that the same branch length in the CC22 tree represented 27 mutations and one recombination event and for CC30 represented 61 mutations and six recombination events. Assuming an equal rate of mutation and recombination in the two CCs, this indicates that CC30 could have more distant common ancestor. CC30 was estimated to have had 1620 recombination events, as opposed to only 240 in CC22, however estimated recombination tract length was on average longer in CC22, at 2173 bases compared to 1694 in CC30 (Table 5.6).

It is notable that the MRSA and MSSA isolates in each CC were completely separated with a more recent common ancestor in the ClonalFrame tree, with the exception of one CC30 isolate, which interestingly represented a CipS case taken from the CipS case-control study in Chapter Three. This finding is consistent with a single emergence of MRSA (and capture of SCC*mec*) underlying many or all of the strains responsible for each of the hospital-associated MRSA lineages in the UK. Further

investigation of the CipS isolate revealed that it contained the *mecA* gene identical to that of all other MRSA isolates in CC30, suggesting its method of meticillin resistance was the same as the other isolates. However, the mapping of the CipS isolate to the SCC*mec* element surrounding the *mecA* gene (not including *mecA*) consisted of 94% unmapped positions, suggesting that the CipS isolate contained an SCC*mec* other than SCC*mec* II found in the MRSA252 reference. Investigation of the SCC*mec* element by Dr Elizabeth Batty using Velvet *de novo* assembly (Zerbino and Birney 2008) revealed high similarity to SCC*mec* V.

5.4.4.2 Within CC evolution

Estimates of ρ and Θ from ClonalFrame suggest ρ/Θ (the ratio of recombination events to mutation events) of 0.05 for CC22 and 0.09 for CC30, implying that for every recombination event there were 20 mutation events in CC22 and 11 mutation events in CC30. In contrast, the estimate of r/m (the ratio of the number of base changes due to recombination to base changes due to mutation) was approximately the same in each CC, in spite of the difference in tract length. Therefore, in CC22 although the average mutation length was 500 bases longer, the same number of bases were changed in each event, perhaps because recombination was more likely to come from other isolates within the CC which is less diverse. Additionally, for each CC the external to internal branch length ratio compared to its coalescent expectation under a constant population size was significantly indicative of clonal expansion (P<0.000005 for both CCs) in fitting with the expansion of the two epidemic CCs.

Table 5.6: Recombination and mutation events inferred from the ClonalFrame tree in CC22 and CC30

	CC22	**CC30**
Number recombination events	249	1620
Number substitutions via recombination	3604	19920
Recombination length	2173 bases	1694 bases
ρ/Θ	0.051	0.092
r/m	0.599	0.920
External/Internal branch length ratio	3.8	3.7

The d_N/d_S ratio of non-synonymous to synonymous mutations was calculated for each CC compared to the reference sequence. Values for both CCs were indicative of purifying selection, with a significantly higher proportion of non-synonymous point mutations in CC22, which had a d_N/d_S ratio of 0.79 compared to 0.58 for CC30 (Exact test P<0.005) (Table 5.7). The higher d_N/d_S for CC22

suggests that although it is largely subject to purifying selection, it has been under an increasing level of directional selection compared to CC30.

Table 5.7: Number of non-synonymous and synonymous mutations in each CC

	CC22	CC30
Number non-synonymous substitutions (d_N)	2960	3378
Number synonymous substitutions (d_S)	1246	1947
d_N/d_S ratio	0.79	0.58

5.4.5 Structure of within CC diversity over space and time

Isolates within each CC were divided into four spatial-temporal groups: Oxford Old, Oxford New, Brighton Old and Brighton New to investigate the diversity within and between each group. Firstly, comparison of average pairwise difference (π_{xy}) over space and time revealed similar differences between isolates separated by geographic space and time (Table 5.8), suggesting that at these scales, space and time have approximately equal effects on the accumulation of diversity. CC30 had greater values for both space and time further supporting its increased diversity.

Table 5.8: Average pairwise nucleotide distance (π_{xy}) for populations separated by space and time for each CC

Average pairwise nucleotide distance (π_x) within each group is represented on the diagonal.

	CC22				CC30			
	Oxford Old	**Oxford New**	**Brighton Old**	**Brighton New**	**Oxford Old**	**Oxford New**	**Brighton Old**	**Brighton New**
Oxford Old	510	617	620	658	985	859	923	869
Oxford New		129	192	172		503	659	500
Brighton Old			245	228			762	668
Brighton New				182				446

The spatial-temporal group with most diversity in both CC22 and CC30 was Oxford Old (π_x=510 for CC22, π_x=985 for CC30) and the group with the least diversity for CC22 was Oxford New (π_x=129) and for CC30 was Brighton New (π_x=446). In both CCs, isolates from each geographic location and the new time period had lower average pairwise differences than isolates from the old time period. Comparisons between groups revealed isolates from the Oxford Old group to be most different from isolates in other groups. However, the distribution of differences between groups may partly be due to

sampling, since isolates from Brighton were from only invasive disease samples, whereas those from Oxford were both carriage and disease. If isolates from Brighton had included carriage samples and those with other disease presentation, their diversity may have increased and Brighton Old isolates may have had similar diversity to those in the Oxford Old group.

Investigation of spatial-temporal groups from the ClonalFrame tree revealed clear evidence for clonal structure, particularly in MRSA isolates from new time groups (Figures 5.7 and 5.8). This was especially the case for CC30 Brighton New MRSA, where all isolates in this group but one formed a distinct cluster. The similarly of isolates from the Brighton New MRSA group is highlighted by the very low π_x value of an average of 31 pairwise differences between isolates in the group, lower than the π_x value of any other spatial-temporal group within CCs separated by meticillin sensitivity, of which the highest was 1127 differences for Brighton New MSSA.

Figure 5.7: ClonalFrame phylogenetic tree for CC22

Terminal branches are coloured according to which spatial-temporal group isolates are from. Dashed red line indicates point at which above all isolates are MRSA and below are MSSA. Scale bar indicates the average number of mutation - recombination events that occur during the interval of time represented by the scale.

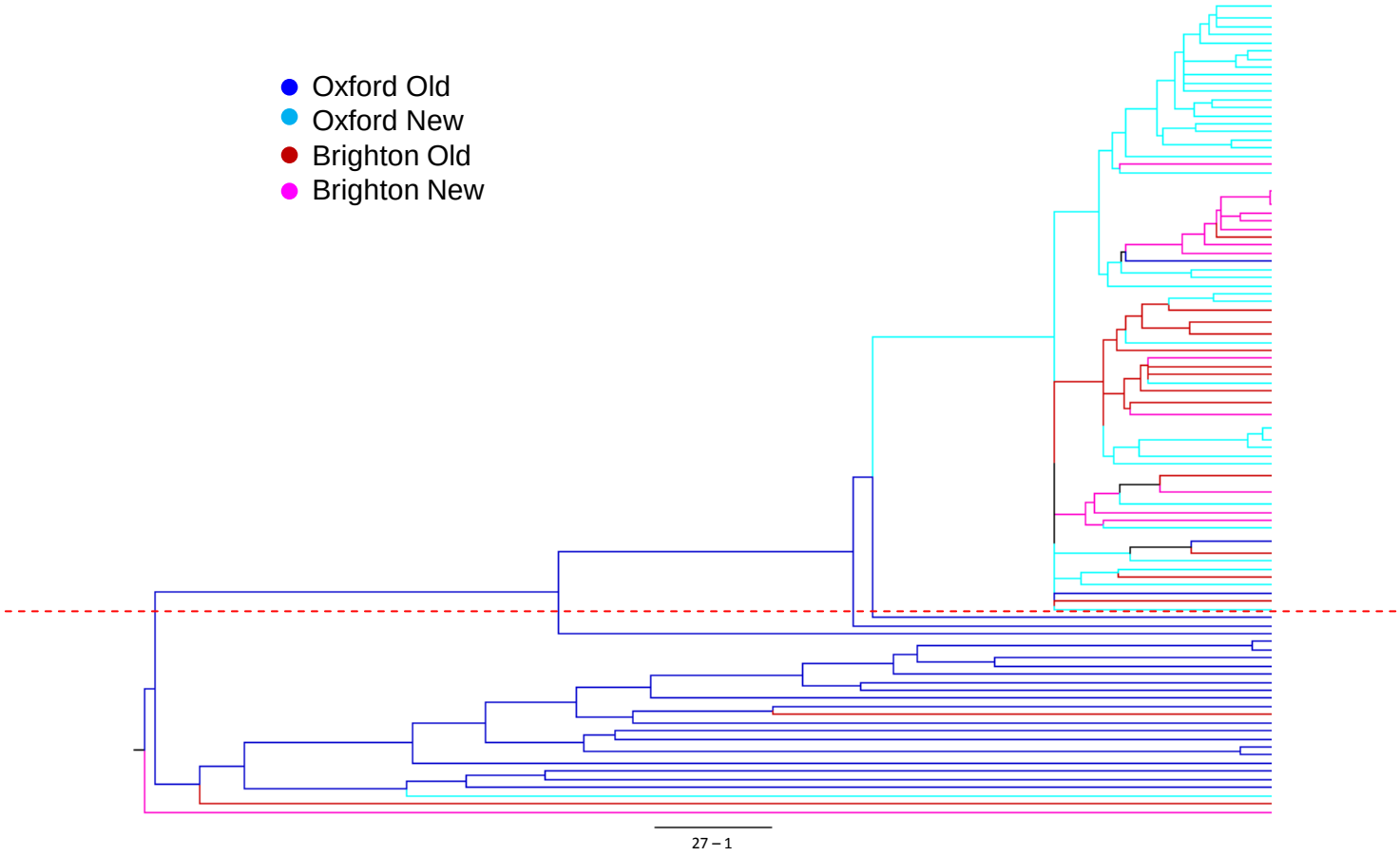
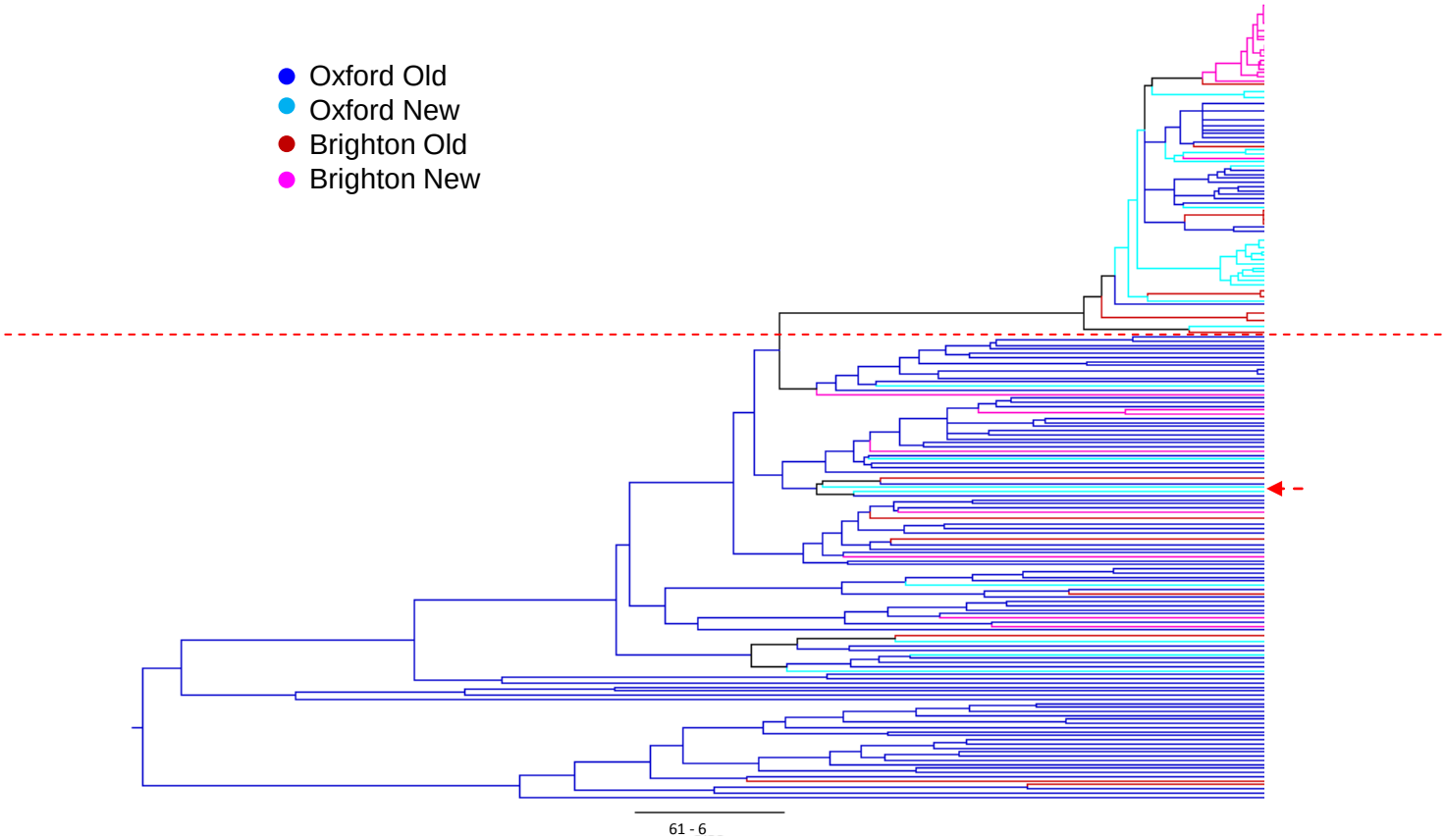


Figure 5.8: ClonalFrame phylogenetic tree for CC30

Terminal branches are coloured according to which spatial-temporal group isolates are from. Dashed red line indicates point at which above all isolates are MRSA and below are MSSA. Dashed red arrow indicates one CipS MRSA isolate amongst MSSA isolates. Scale bar indicates the average number of mutation - recombination events that occur during the interval of time represented by the scale.



5.4.5.1 Additional spatial-temporal structure within CC structure from the pan-genome

To represent the additional diversity found in contigs from the non-reference pan-genome, a matrix was constructed of presence or absence of each contig in each isolate. Isolates in the matrix were ordered the same as the ClonalFrame tree to allow comparison of findings from the pan-genome with mapping to the CC-specific references. Alignment of the ClonalFrame tree and the matrix of contigs from the non-reference pan-genome is shown in figures 5.9 and 5.10.

Figure 5.9: Congruence of ClonalFrame tree and matrix of non-reference contigs from the pan-genome in CC22

The matrix represents each isolate in the ClonalFrame tree on the y axis, and the x axis is each contig found in at least one isolate in that CC, in order of contigs in the pan-genome. Black indicates presence of a contig, and white absence. Thin arrow indicates the most common additional gene identified from the pan-genome. Thick arrows indicate blocks of contigs which are phages.

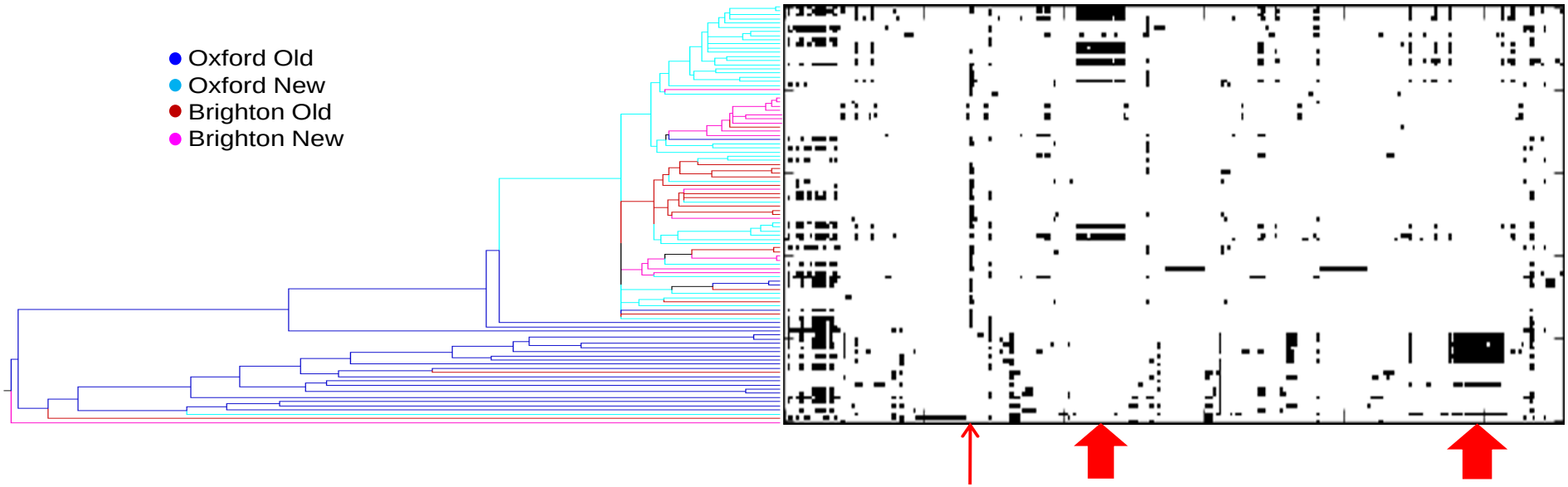
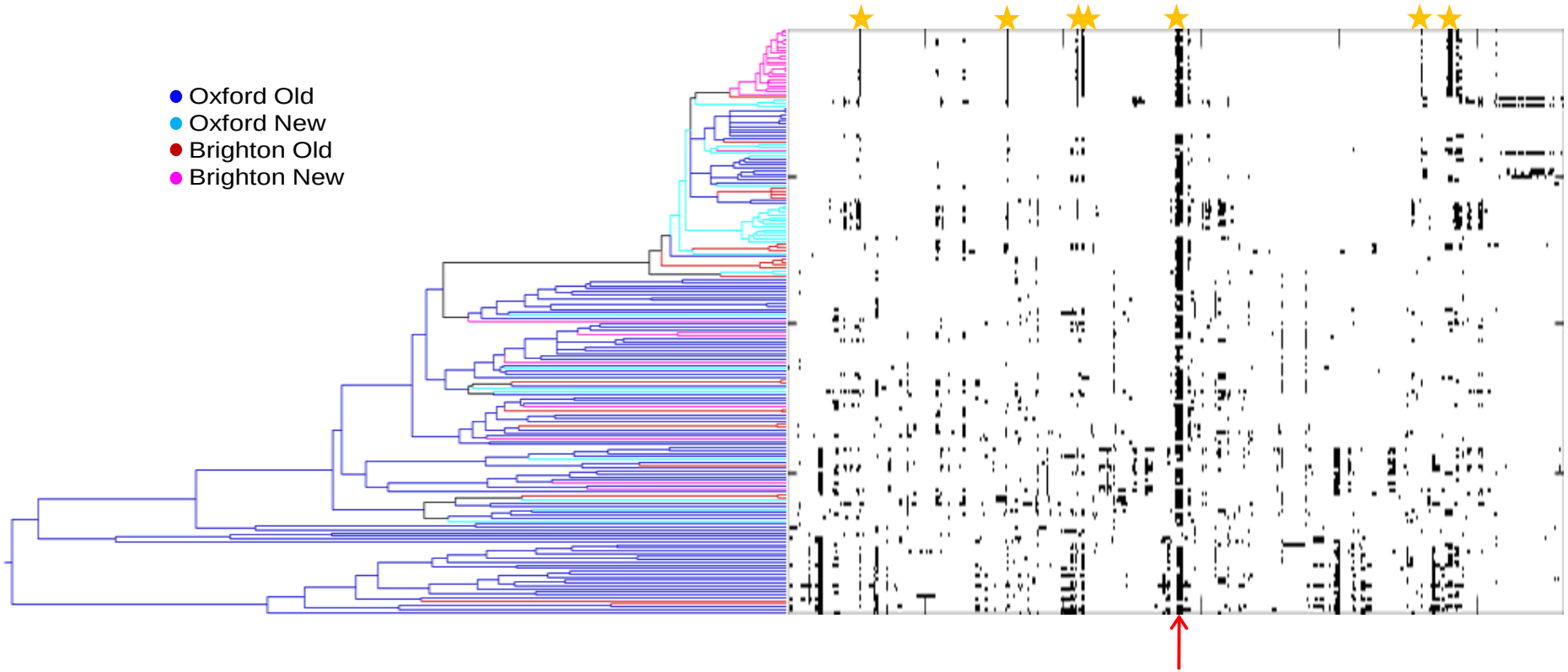


Figure 5.10: Congruence of ClonalFrame tree and matrix of non-reference regions from the pan-genome in CC30

The matrix represents each isolate in the ClonalFrame tree on the y axis, and the x axis is each contig found in at least one isolate in that CC, in order of contigs in the pan-genome. Black indicates presence of a contig, and white absence. Stars are contigs associated with the cluster of MRSA isolates from the Brighton New period. Arrow indicates the most common additional gene identified from the pan-genome.



There was no clear relationship between the matrix of contigs from the non-reference pan-genome and the ClonalFrame tree. However, there were some isolated regions of correlation, for example in CC22 the phages highlighted in figure 5.9 were present only in isolates from the same spatial-temporal group; and in CC30 there were some regions in the pan-genome that mapped predominantly to the Brighton New MRSA isolates, which also group together in the ClonalFrame tree (Figure 5.10). Nonetheless, the majority of diversity identified from the pan-genome was different from that identified when mapping to a reference consisting largely of the core genome. This lack of congruence is perhaps because mobile elements are inserted and excised multiple times during the history of a CC; and because the methodology does not distinguish similar but actually different versions of the same genes.

To further investigate whether the non-reference pan-genome revealed any structure correlating to spatial-temporal groups within CCs an association style study was carried out on the genes from the non-reference pan-genome. For this, the number of isolates in each spatial-temporal group that contained each gene was counted and a Fisher's Exact Test performed to determine whether genes were over-represented in particular groups. Due to the large number of comparisons performed, *P* values were corrected using the Bonferroni correction, whereby the *P* value required to obtain significance (in this case *P*=0.01) was divided by the number of tests performed (279 for CC22 and 564 for CC30).

In CC22 12 genes were distributed significantly differently between groups. The majority (7/12) were found more in the Oxford Old group than any other. None of the genes had associations with virulence according to an NCBI PubMed literature search (http://www.ncbi.nlm.nih.gov/pubmed) of the Basys annotations. Similarly in CC30 26 genes were distributed significantly differently between groups including 10 genes found more in the Brighton New group and 11 found less in the Oxford Old group. The regions found in CC30 Brighton New isolates were of particular interest, since this was the only group that showed clear congruence with the ClonalFrame tree (Figure 5.10), and included four subsequent genes consisting of a phage, an autolysin and the Tn917 transposon. Interestingly 4/11 genes found less in the Oxford Old group had a putative association with virulence according to a NCBI PubMed search, including the toxic-shock-syndrome toxin (Tristan, Ferry et al. 2007) and the putative adhesin *sdrD* (Peacock, Moore et al. 2002). This suggests that on average isolates from the

CC30 Oxford Old group may have reduced virulence, perhaps due to the nature of the Oxford Old sample set which consisted of 49% carriage samples.

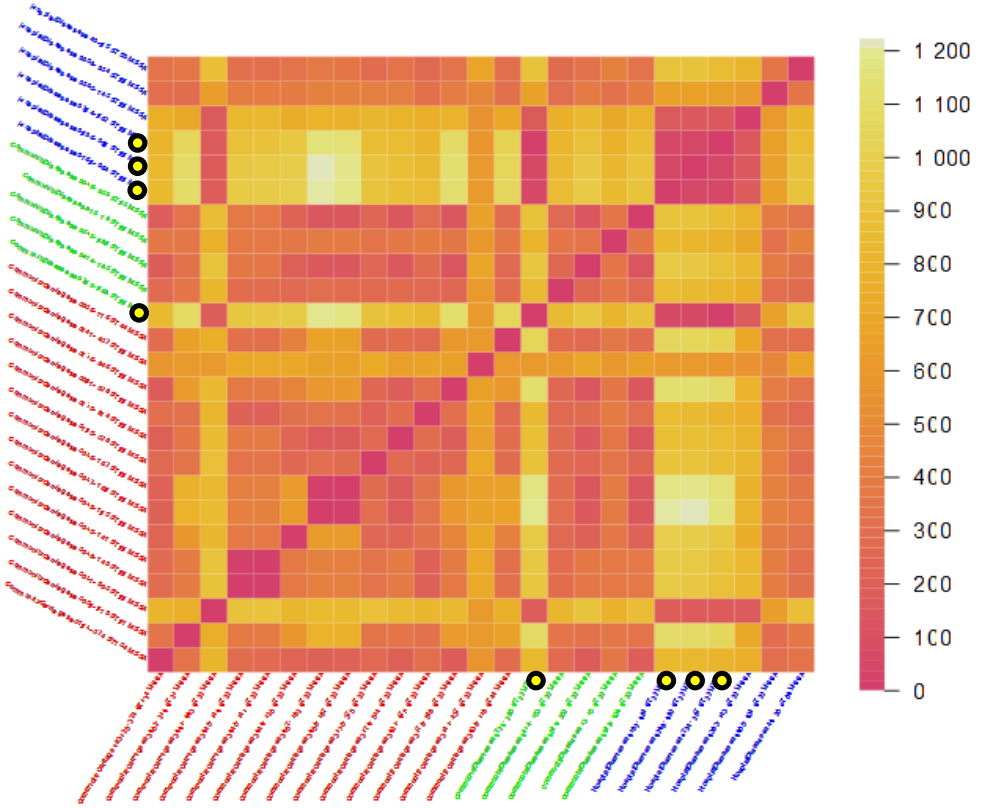
5.4.6 Colonising versus infecting isolates

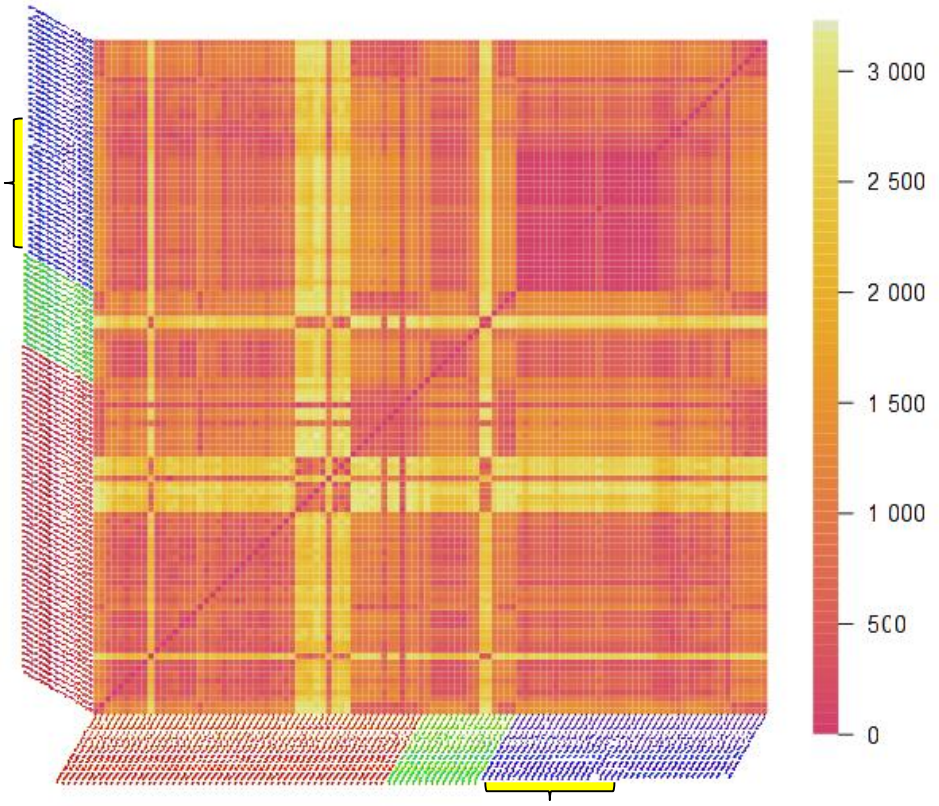
Investigation of colonising versus disease isolates involved only isolates from the Oxford Old group, since samples in this group were initially collected to compare carriage with community and hospital-associated disease (where community-associated was defined as isolation within 24 hours of hospital admission). Previous investigation of this dataset using PCR found seven genes associated with virulence (Peacock, Moore et al. 2002) (section 1.3.1), however the high resolution of whole-genome sequencing may allow identification of further differences between samples from the Oxford Old group.

To investigate differences between carriage and disease isolates from the Oxford Old dataset heat maps were drawn to depict the diversity between carried, community-associated disease and hospital-associated disease isolates (Figure 5.11). The heat maps did not reveal that isolates from the same disease group were more closely related to one another than isolates from different groups.

Fig 5.11: Heat map of variant positions between every Oxford Old isolate in CC22 (left) and CC30 (right)

Each square represents the number of variant positions between two isolates on a colour scale as represented to the right of each heat map (note the greater scale in CC30). Isolates are ordered according to carriage and disease status. Blue legends are hospital disease, Green community disease and Red carriage. Yellow dots and brackets mark isolates that are meticillin resistant.





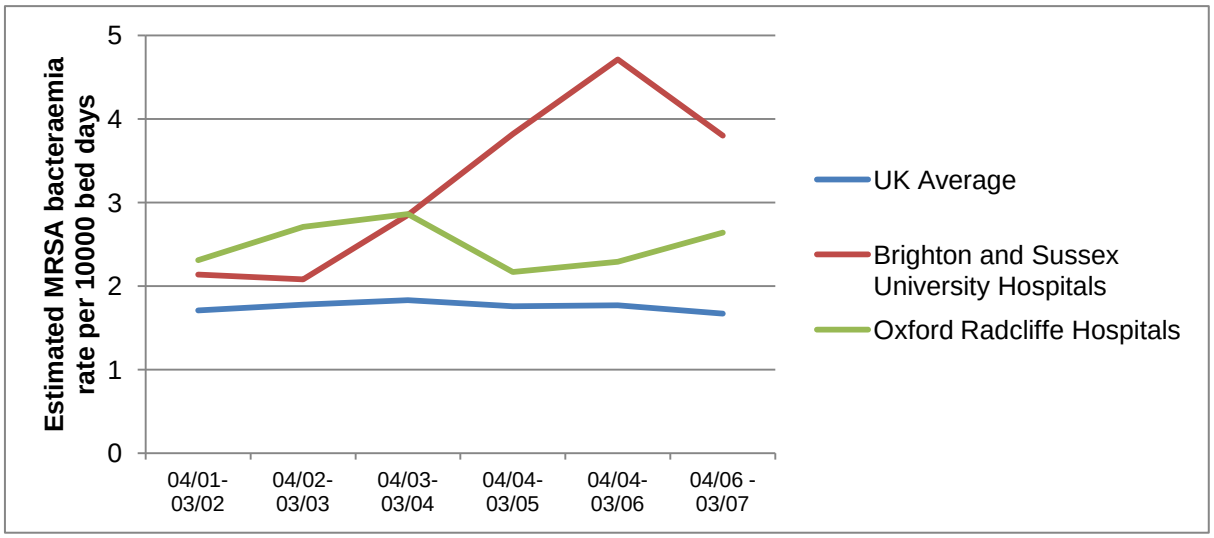
To further investigate the diversity between the carried, community-associated disease and hospital-associated disease isolates from the Oxford Old group the distribution of the genes in the non-reference pan-genome was considered. Here the number of isolates from each of the three groups that contained each gene was counted and a Fisher's Exact Test performed, with *P* values Bonferroni corrected, to determine whether genes were over-represented in any group. No genes differed significantly between the three groups.

5.4.7 CC30 Brighton New MRSA cluster

Investigation of spatial-temporal diversity within each CC identified one group of particularly similar isolates within CC30: the Brighton New MRSA cluster, a group of isolates (all ST36) that were distinct from other isolates in that CC. In the ClonalFrame tree every Brighton New MRSA isolate but one grouped together in a separate distinct cluster; and in the pan-genome Brighton New isolates shared many genes and showed congruence with the ClonalFrame tree unlike that seen for any other isolates, and likely reflecting their recent common ancestor (Figure 5.10).

Further investigation revealed that between 2006 and 2007, the period the New Brighton samples were collected, Brighton was experiencing unusually high and unexplained MRSA rates (Figure 5.12). This led to generation of the hypothesis that: the CC30 MRSA Brighton New cluster represents a genetic explanation for the MRSA epidemic in Brighton between 2006-07.

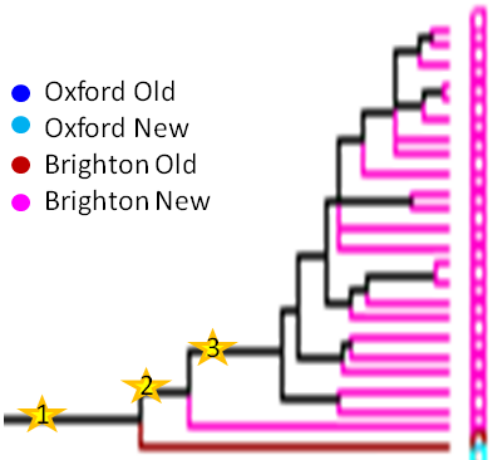
Figure 5.12: Estimated MRSA bacteraemia rate per 10000 bed days in Brighton, Oxford and National Average between April 2001 and March 2007 (Health Protection Agency 2009)



Investigation of the CC30 Brighton New MRSA cluster from the ClonalFrame tree revealed that on the three branches leading to the cluster there were 29 mutations of which 24 were non-synonymous, and no recombination events (Figure 5.13). A search for each gene containing a non-synonymous mutation in the NCBI PubMed database revealed one non-synonymous mutation in a gene with a potential virulence association encoding a fibronectin-binding protein (Peacock, Moore et al. 2002), located on branch three (Figure 5.13). The d_N/d_S ratio on the three branches leading to this cluster was 1.6, which represented a significantly larger proportion of non-synonymous mutations than seen in the rest of the CC (P=0.02), suggesting that this cluster was under directional selection. Additionally, the external to internal branch length ratio compared to the coalescent expectation for a constant population size was 2.5, which although lower than those seen for the entire CCs, likely due to the smaller sample size of this group, still provided significant support (P=0.00006) for clonal expansion.

Figure 5.13: Branches on the CC30 ClonalFrame tree leading to the Brighton New MRSA cluster

Branch	Number mutations	Number non-synonymous
1	12	7
2	8	8
3	9	9



Presence of additional genes in the accessory genome of the CC30 Brighton New cluster was investigated firstly by identifying genes found significantly more in the Brighton New isolates than any other spatial-temporal group and secondly by investigating all genes that were shared between the Brighton New MRSA isolates. This resulted in association of 13 genes from the non-reference pan-genome with the Brighton New MRSA cluster, eight of which were found using both methods. A literature search using NCBI PubMed revealed two of the genes had putative virulence associations: the putative adhesin *sdrD* (Peacock, Moore et al. 2002), and a superantigen-encoding pathogenicity island protein (Novick 2003). Therefore, combined with the results from the CC-specific reference, three genes associated with virulence were associated with the Brighton New MRSA cluster, suggesting that it could cause more severe disease than the other isolates.

5.5 Discussion

Investigation of ST22 and ST36 using whole-genome sequencing supports the hypothesis that sequence diversity exists within populations of a single *S. aureus* MLST, since both STs had a mean pairwise difference of over 90 variant positions. Additionally, only two pairs of indistinguishable isolates were identified throughout all 300 isolates, revealing that the majority of *S. aureus* isolates are unique at the whole-genome level.

Comparison of diversity between isolates in CC22 and CC30 falsified the second hypothesis that CC22 and CC30 have similar levels of diversity. CC30 was shown to have greater diversity than

CC22 with a π_x of 853 compared to only 367 for CC22, depiction of the diversity in a heat map (Figure 5.3), and the ClonalFrame trees (Figures 5.7 and 5.8) supported this finding. Finally, investigation of spatial-temporal variation within each CC revealed a clear clonal structure over space and time, as demonstrated by the ClonalFrame trees (Figures 5.7 and 5.8), in line with the final hypothesis that CC22 and CC30 sequence diversity is spatially and temporally structured.

Although the finding of within ST diversity is no longer novel, it was unknown at the time of designing this study, and now adds important support to other work. Whole-genome sequencing of 63 isolates with ST239 revealed 4310 variant positions in the core-genome alone (Harris, Feil et al. 2010), a number greater than found in either ST22 or ST36 in this study, despite a smaller sample size. This may reflect the increased geographic diversity of the ST239 which were from a global collection.

This study was the first to investigate the whole genomes of *S. aureus* isolates from an entire CC. Use of two different CCs, and meticillin-sensitive as well as resistant samples from a variety of disease presentations, means that findings are more applicable to *S. aureus* as a whole compared to more specific investigations of MRSA samples within one ST. As well as revealing the greater diversity of CC30, it was also possible to investigate the mechanisms responsible for the maintenance of diversity in the two CCs, which were similar. Within each CC mutation occurred at least 10 times more often than recombination events, and r/m values were comparable to previous work using MLST data (Feil, Cooper et al. 2003; Vos and Didelot 2009). Therefore, whole-genome sequencing from this study verifies the use of MLST to quantify the impact of recombination on sequence diversity. Additionally, as supported by previous investigation, MRSA isolates in each CC were much less diverse, than MSSA isolates (Enright, Day et al. 2000; Grundmann, Aanensen et al. 2010).

The study was also able to reveal a structured relationship between isolates from different geographic regions and time periods. Comparison of diversity between isolates separated over space and time revealed that a time period of 10 years has a similar effect on diversity to geographic separation in two cities nearly 100 miles apart, a novel finding, suggesting that *S. aureus* does transmit between hospitals, but at a slow rate.

As opposed to the clear structure seen in the sequence diversity over space and time, there was no evidence for structure within CCs correlating with disease phenotype. Since, Peacock *et al* did find significant association of seven genes with virulence, it is likely that if all the samples from the Oxford Old set were investigated, rather than only the ones in CC22 and CC30, then significant differences between disease groups would be identified. However, since this study was not specifically designed to identify differences between carried and disease isolates it is unsurprising that Peacock *et al*'s results cannot be replicated.

Finally, investigation of spatial-temporal variation within CCs led to identification of a group of isolates that were particularly similar to others within their CC. This allowed generation of a further hypothesis, that the cluster may represent a genetic explanation of an MRSA outbreak in Brighton. Investigation of this hypothesis revealed the cluster had a strong suggestion of population expansion and directional selection, in fitting with epidemic properties. An epidemic explanation for the outbreak is likely above an environmental explanation such as bad hospital hygiene, since environmental factors would have caused all *S. aureus* strains to increase in frequency, rather than a single cluster of closely related isolates. However the possibility of a "super-spreader", for example a single individual transmitting the same strain of *S. aureus* to many people, cannot be excluded, since it might also produce a cluster with features of a clonal expansion.

5.5.1 Limitations and compromises

The primary and unavoidable limitation of the study was the use of retrospective isolates from pre-defined collections. Each spatial-temporal group was taken from a different collection made to investigate separate hypotheses. Therefore, the samples were not distributed evenly with respect to disease presentation and place of onset. However, it is unlikely that this had an effect on the associations found, since investigation of the Oxford Old set, which includes carriage, community and hospital acquired disease isolates, showed no difference in diversity between samples with different disease presentations and place of onset (section 5.4.6).

As all of the samples from Brighton were hospital acquired disease, it could similarly be argued that the CC30 Brighton New cluster was in part found due to the initial similarity of isolates in this group compared to those from Oxford. However, if this was the causal factor for the CC30

Brighton New MRSA cluster, an equally high level of similarity would have been expected between isolates from the CC30 Brighton Old, and both CC22 Brighton groups, which was not the case. Therefore, it is unlikely that the epidemiological differences between samples in each CC had an impact on the relationships seen between isolates separated over space and time.

A further potential limitation of the study arises from sequence assembly by mapping to a CC-specific reference, containing both core and non-core genomic regions. This choice was made in order to maximise the proportion of the sample sequenced that could be accounted for in a simple analytical process, in view of the likelihood that genomes from related lineages share a larger proportion of genes than those chosen at random. This prediction was borne out to a large degree by the observation that 97.9% and 97.2% of the reads from each sample in CC22 and CC30 respectively were mapped (Tables 5.3 and 5.4). The effect of including some MGEs within the CC-specific reference is unclear, but might lead to some confusion in interpreting mapping results.

One area in which the use of an MGE-containing reference might affect inference is in the calculation of measures of molecular diversity such as π_x and π_{xy}. For example, it is conceivable that diversity estimates between genomes that share the presence of diverse copies of an MGE would be inflated compared with comparisons where one of the genomes did not contain an MGE copy, and so such positions would be omitted from the calculation. For this to be a major problem affecting the high-level comparisons made in this study, polymorphic MGEs would have to comprise a large proportion of the genome and also have very different diversity from non-mobile genetic elements, and there is no evidence that this is the case. For example, alignment of the 23 reference genomes available on NCBI using Mauve revealed that 83% of the genomes are shared between all 23 sequences, a proportion that is certain to be higher for within-CC comparisons.

Additionally, inclusion of MGEs in the reference genome may alter inferences made about evolution. It has recently been shown that the non-core genome, containing MGEs, has a significant enrichment of synonymous changes (Castillo-Ramirez, Harris et al. 2011), suggesting that d_N/d_S ratios calculated from data generated using references containing non-core regions may be underestimated. An alternate point of view is that at short timescales, the most

appropriate assessment of evolutionary processes (as summarised by d_N/d_S) should include as much of the genome shared by the samples as possible.

If a different representative of the same CC were used as reference, two kinds of change in the results might be expected. Firstly, the exact proportion of the genome for which sequence calls could be made would be expected to change. However, empirical evidence suggests that this effect is small, because the proportion of the genome lost due to increasing the distance from the reference is substantially smaller than that filtered out due to non-uniqueness. Secondly, the exact compliment of MGEs in the chosen reference will have a small but measurable impact on diversity statistics (as above). These effects could either be adjusted for by post-hoc filtering or quantified by mapping a subset of samples to a variety of references.

5.5.2 Impact of results

Confirmation of within ST diversity suggests it would be possible to use whole-genome sequencing to investigate transmission of common previously indistinguishable lineages, such as investigated in this study, in order to track their evolution and spread around hospitals. This would enable identification of mechanisms through which *S. aureus* is spreading and have implications for infection control interventions.

The within CC clonal structure over space and time reveals a slow rate of *S. aureus* transmission between hospitals, since the diversity seen between two hospitals was similar to that generated within one hospital over a 10 year period. This suggests that *S. aureus* lineages are not confined to a particular hospital, however will not spread rapidly. Therefore if one hospital has a *S. aureus* outbreak, it is important to monitor *S. aureus* strains in other hospitals for characteristics seen in the outbreak, so that the strains involved, and could be prevented from spreading to cause outbreaks in further hospitals.

Finally, the identification of an expanding cluster of isolates in Brighton highlights the potential use of whole-genome sequencing for identifying epidemics, which may appear as groups of even more closely related isolates. In the ClonalFrame tree, the isolate most closely related to the Brighton New MRSA cluster was from the Brighton Old group, and could perhaps represent an early ancestor of the epidemic strain before its expansion. Therefore high resolution investigation

at this time may have identified the lineage when it was beginning to be seen at increased frequency, and initiated preventative action to reduce the extent of the epidemic.

5.5.3 Future work

With next-generation sequencing continuing to increase in speed and decrease in cost, and the possibility of third generation sequencing on the horizon (Travers, Chin et al. 2010) results here represent only the tip of the iceberg of what whole-genome sequencing may be able to offer in future. In order to generalise further from this study it would be interesting to sequence samples from other *S. aureus* CCs in order to investigate hypotheses such as: UK epidemic *S. aureus* CCs 22 and 30 have lower diversity than other non-hospital associated CCs. This would be particularly interesting if CCs predominantly found in the community, or associated with carriage, such as CC45 (Melles, Gorkink et al. 2004) were used. Additionally, whole-genome sequencing of isolates from other CCs may be able to identify and prevent epidemics arising from clonal expansions of "wild" strains in the community (as suggested in Chapter Three). This would enable hypotheses such as: whole-genome sequencing can identify clonal expansions of *S. aureus* in the community, to be answered.

Furthermore, it is possible that the maximum resolution offered by whole-genome sequencing will enable identification of previously unrecognised differences between carriage and disease causing isolates. Comparison of a large group of carriage and disease strains using whole-genome sequencing could enable an association study to investigate genes causing particular disease phenotypes. In particular, sequencing of the entire Oxford Old sample set would allow comparison between carriage; and hospital and community onset disease isolates, enabling investigation of hypotheses such as: carried *S. aureus* isolates differ genetically from those causing disease.

6 Discussion

6.1 Summary of key findings

This thesis has investigated the *S. aureus* host-organism relationship in depth, focussing particularly on the under-investigated areas of *S. aureus* in the community; and population genetic features of epidemic *S. aureus* lineages, investigated using whole-genome sequence analysis. Results from the thesis provide new and currently relevant knowledge of the interaction between *S. aureus* and its host.

In Chapter Three, individuals with MRSA and no traditional nosocomial risk factors, namely those with CA-MRSA, were actively searched for and compared to those with healthcare-acquired MRSA. Eleven of 6801 individuals investigated were found to have convincing CA-MRSA, after using a combination of definitions requiring investigation of host factors, such as previous hospital exposure, and organism factors, particularly ciprofloxacin sensitivity. It is therefore clear that CA-MRSA levels in Oxfordshire are low, and rigorous definitions are needed for reliable identification. Fifteen further individuals were identified who fit the host-based CDC definition for CA-MRSA but had nosocomial-type *S. aureus* isolates. These strains were termed “feral” MRSA, since they represented hospital strains transmitting in the community.

Chapter Four documented the design and results from the biggest *S. aureus* carriage study to date. This study investigated host demographic and healthcare risk factors along with organism factors following *S. aureus* carriage and how it changed over time. The study revealed male sex, current employment, participation in contact sport and not having a recent out-patient appointment as risk factors for carriage of *S. aureus* in the healthy community. All factors but male sex had not previously been identified as risk factors for *S. aureus* carriage, highlighting the benefits of a large study investigating comprehensive risk factors. Organism-level analysis showed that median *S. aureus* carriage durations are longer than one year and that 65% of people initially identified as carriers carry persistently, of whom 26% carry more than one type of *S. aureus*. In the remaining 45% carriage in the nose appears to be a dynamic phenomenon changing over time.

A joint analysis of host and organism factors showed that younger individuals were more likely to carry *S. aureus* transiently, with strains changing over time; and had increased chance of identification of *S. aureus* from later swabs if they began the study negative. Investigation of samples from the two predominant UK healthcare associated lineages (CC22 and CC30) revealed that they were more likely to be carried singly and persistently, perhaps suggesting they are able to out-compete other strains, thus offering a potential explanation for their success.

Finally, in Chapter Five whole-genome sequencing was used to distinguish between apparently identical isolates within CC22 and CC30. Although this has since been demonstrated (Harris, Feil et al. 2010), when the study was designed it was not yet formally known whether common strains indistinguishable by MLST would differ even at the whole-genome level. Isolates in CC22 and CC30 revealed a clear clonal structure over space and time, and showed that a period of 10 years and geographic space of 100 miles result in equal levels of *S. aureus* diversity. Identification of an even more closely related cluster of isolates within CC30 correlated with an epidemic in Brighton, highlighting that whole-genome sequencing could both identify an epidemic expansion of *S. aureus* and elucidate possible explanatory genetic factors associated with such spread.

6.2 Contributions to the field

Investigation of CA-MRSA undertaken for this thesis was more comprehensive than any previous investigation of CA-MRSA in the UK, in that it used large sample sets and tested multiple definitions. Rates of reliably identified CA-MRSA were low (0.2%), with an additional 0.2% of feral strains, not previously identified in the UK, found circulating in the community, both in the studies described in Chapter Three and in the carriage study in Chapter Four. That CA-MRSA is low in the UK provides confirmation that an epidemic such as seen in the USA, where CA-MRSA (mainly USA300) makes up 59% of emergency unit admissions for skin infection (Moran, Krishnadasan et al. 2006), is currently unlikely. Additionally, Chapter Four showed that *S. aureus* carriage duration often lasts longer than one year, and common healthcare associated strains are more likely to be carried as a single persistent strain. So, although levels of CA-MRSA are low, research for this thesis has revealed that people without recent hospital contact may still carry hospital acquired *S. aureus* strains long after they acquired them, which might explain the presence of feral strains in the community.

The carriage study in Chapter Four represented the largest study of *S. aureus* carriage, including collection of detailed host and organism factors, to date. Investigation of individuals from the community with no known underlying disease revealed two new risk factors associated with *S. aureus* carriage: participation in contact sport and current employment, and a protective effect of a recent out-patient appointment. It is likely that sport and employment represent interplay between younger age and close contact with other individuals, increasing the chances of acquisition of *S. aureus*. The relationship found between younger age and transient carriage has also never been described before, probably because the study included a unique combination of host and organism factors never before considered together. The carriage study also investigated a large group of individuals who were initially carriage-negative for acquisition of carriage, of whom 10% had *S. aureus* isolated from at least one subsequent swab within eight months. *S. aureus* acquisition has never been investigated at such length in individuals without predisposing risk factors, and as the study continues it will be able to give a more comprehensive estimation of carriage acquisition.

This thesis represents the first study of large numbers of isolates from the same *S. aureus* CC, rather than just a single ST, undertaken by next-generation sequencing. It also describes one of the largest collections of *S. aureus* to be whole-genome sequenced, analysed and reported. Firstly, the filters developed to produce reliable sequence data after assembly by mapping to a reference represent a practical addition to the field that will be useful for future studies using the same technique. Results from whole-genome sequencing revealed information about the effects of time and geography on variation amongst strains from the same CC. The study demonstrated that whole-genome sequencing can be used to identify outbreaks, defined as groups of highly similar strains sharing space and time. More detailed analysis of gene content provided insights about genes that may account for increases in epidemic spread of a lineage. This suggests a new way to identify potential novel variants with epidemic potential. These features may offer substantial benefit in ongoing surveillance, monitoring for the emergence of hyper-virulent and transmissible variants that are infection threats.

6.3 Future work

The carriage study designed in Chapter Four provides a multitude of opportunities for future research. As the study continues to progress for up to four years, further inferences about

carriage duration and how it changes over time will be possible. Additionally, revisiting GP records after two (and four) years will allow correlation of host healthcare behaviour with carriage over time, something never previously documented in *S. aureus*. Further arms of the carriage study, not analysed for this thesis, include recruitment of individuals from hospital likely to have increased co-morbidities and recruitment of participants with clinical samples, both from invasive and superficial infections. This will allow comparison of host risk factors for different *S. aureus* presentation, as well as comparison of organism factors between strains with different manifestations.

The carriage study was also the first study to regularly investigate carriage behaviour over a two year period. It revealed that *S. aureus* carriage is dynamic and may involve evolution of strains carried in the nose over time as well as strain acquisition and loss. Statistical modelling of *S. aureus* carriage over time would provide a way to investigate *S. aureus* carriage and incorporate all the variation seen in carriage behaviour, and strain change by evolution or acquisition and loss. Such a model might be able to relate further host factors to *S. aureus* carriage dynamics, and perhaps give additional insight into the correlation between age and type of *S. aureus* carriage. Since *S. aureus* carriage is a risk factor for disease (Corbella, Dominguez et al. 1997; von Eiff, Becker et al. 2001; Wertheim, Vos et al. 2004; Safdar and Bradley 2008), it would also be interesting to investigate whether different carriage behaviour over time correlates with likelihood of progression to disease, perhaps with specific presentations.

This study collected more detailed risk factor information for *S. aureus* carriage than ever before and identified new correlates with *S. aureus* carriage and its behaviour over time. Despite this, the risk factors found did not completely explain the extent of variation seen in *S. aureus* carriage behaviour. Expansion of the study to investigate two further factors thought to affect *S. aureus* carriage: host genetics and colonisation with other bacteria, may help to further explain *S. aureus* carriage and how it changes over time. An application for ethical approval for collection of human DNA as part of this study is currently being submitted. Addition of human genetic data to the information already collected may reveal factors correlated with presence or absence of *S. aureus* carriage as well as how it changes over time. The nasal swabs returned as part of the study could also be used to investigate other bacteria present in the nose. Although this is not part of the

routine work currently performed on arrival of nasal swabs, as the study continues to four years, it would be possible to alter the protocol to investigate other bacteria.

Combination of results from the carriage study with whole-genome sequencing, as seen in Chapter Five, would produce an exemplary rich dataset which could be used to answer many questions surrounding *S. aureus*. One hundred randomly selected isolates from the study have already been completely sequenced and will enable detailed investigation of the diversity of *S. aureus* carriage in Oxfordshire. This sampling frame is an unbiased collection of community strains, and is therefore unlikely to have any underlying structure that may confound results. The approach taken in Chapter Five, expanded to a larger sampling frame over time, will provide new understanding of *S. aureus* evolution. Additionally, the very large collection of *S. aureus* being assembled in the carriage studies could be used in case-control genome wide studies of bacterial factors that associate with invasive disease, virulence, or even duration and persistence of carriage.

Whole-genome sequencing would also allow investigation of evolution and change of strains in the nose over time at considerably higher resolution than *spa* typing. This would reveal how strains in single-persistent carriers change over time, and enable calculation of the rate of *S. aureus* evolution in the nose. Subsequent isolates have already been sequenced from one participant who progressed to disease during the study, whilst carrying the same *spa* type throughout; this identified a series of mutations differing between the carried and disease causing samples. These mutations will be further investigated to identify possible disease causing genes, which could become future targets for *S. aureus* therapy.

The carriage study set up as part of this thesis, combined with rapidly advancing technology allowing fast and increasingly affordable whole-genome sequencing; both of bacteria, as demonstrated in this thesis, and of humans; provides an opportunity to further investigate the *S. aureus* host-organism relationship. The work presented in this thesis provides the foundation for a series of investigations into *S. aureus* carriage dynamics and associated host factors. Such investigations will use data from the carriage study, and incorporate other findings from this thesis, and may further explain why some individuals suffer from *S. aureus* disease, eventually leading to new ways to reduce the *S. aureus* disease burden.

7 Literature cited

Abudu, L., I. Blair, et al. (2001). "Methicillin-resistant Staphylococcus aureus (MRSA): a community-based prevalence survey." Epidemiol Infect 126(3): 351-356.

Adedeji, A., T. M. Weller, et al. (2007). "MRSA in children presenting to hospitals in Birmingham, UK." J Hosp Infect 65(1): 29-34.

Aires de Sousa, M. and H. de Lencastre (2003). "Evolution of sporadic isolates of methicillin-resistant Staphylococcus aureus (MRSA) in hospitals and their similarities to isolates of community-acquired MRSA." J Clin Microbiol 41(8): 3806-3815.

Aires de Sousa, M. and H. de Lencastre (2004). "Bridges from hospitals to the laboratory: genetic portraits of methicillin-resistant Staphylococcus aureus clones." FEMS Immunol Med Microbiol 40(2): 101-111.

Armstrong-Esther, C. A. (1976). "Carriage patterns of Staphylococcus aureus in a healthy non-hospital population of adults and children." Ann Hum Biol 3(3): 221-227.

Atkinson, S. R., J. Paul, et al. (2009). "The emergence of meticillin-resistant Staphylococcus aureus among injecting drug users." J Infect.

Berglund, C., P. Molling, et al. (2005). "Multilocus sequence typing of methicillin-resistant Staphylococcus aureus from an area of low endemicity by real-time PCR." J Clin Microbiol 43(9): 4448-4454.

Bhattacharya, D., H. Carleton, et al. (2007). "Differences in clinical and molecular characteristics of skin and soft tissue methicillin-resistant Staphylococcus aureus isolates between two hospitals in Northern California." J Clin Microbiol 45(6): 1798-1803.

Bischoff, W. E., M. L. Wallis, et al. (2004). "Staphylococcus aureus nasal carriage in a student community: prevalence, clonal relationships, and risk factors." Infect Control Hosp Epidemiol 25(6): 485-491.

Boye, K. and H. Westh (2011). "Variations in spa types found in consecutive MRSA isolates from the same patients." FEMS Microbiol Lett 314(1): 101-105.

Bygott, J., D. A. Enoch, et al. (2008). "Presumed community-acquired meticillin-resistant Staphylococcus aureus (MRSA) isolates reflect spillover of healthcare-associated MRSA." J Hosp Infect 69(2): 197-198.

Castillo-Ramirez, S., S. R. Harris, et al. (2011). "The Impact of Recombination on dN/dS within Recently Emerged Bacterial Clones." PLoS Pathog 7(7): e1002129.

Cespedes, C., B. Said-Salim, et al. (2005). "The clonality of Staphylococcus aureus nasal carriage." J Infect Dis 191(3): 444-452.

Chambers, H. F. (2001). "The changing epidemiology of Staphylococcus aureus?" Emerg Infect Dis 7(2): 178-182.

Chambers, H. F. and F. R. Deleo (2009). "Waves of resistance: Staphylococcus aureus in the antibiotic era." Nat Rev Microbiol 7(9): 629-641.

Charlebois, E. D., F. Perdreau-Remington, et al. (2004). "Origins of community strains of methicillin-resistant Staphylococcus aureus." Clin Infect Dis 39(1): 47-54.

Chin, C. S., J. Sorenson, et al. (2011). "The origin of the Haitian cholera outbreak strain." N Engl J Med 364(1): 33-42.

Conceicao, T., M. Aires-de-Sousa, et al. (2007). "Replacement of methicillin-resistant Staphylococcus aureus clones in Hungary over time: a 10-year surveillance study." Clin Microbiol Infect 13(10): 971-979.

Cookson, B. D., D. A. Robinson, et al. (2007). "Evaluation of molecular typing methods in characterizing a European collection of epidemic methicillin-resistant Staphylococcus aureus strains: the HARMONY collection." J Clin Microbiol 45(6): 1830-1837.

Cooper, J. E. and E. J. Feil (2006). "The phylogeny of Staphylococcus aureus - which genes make the best intra-species markers?" Microbiology 152(Pt 5): 1297-1305.

Corbella, X., M. A. Dominguez, et al. (1997). "Staphylococcus aureus nasal carriage as a marker for subsequent staphylococcal infections in intensive care unit patients." Eur J Clin Microbiol Infect Dis 16(5): 351-357.

Cordova, S. P., C. H. Heath, et al. (2004). "Methicillin-resistant Staphylococcus aureus bacteraemia in Western Australian teaching hospitals, 1997-1999: risk factors, outcomes and implications for management." J Hosp Infect 56(1): 22-28.

Cosgrove, S. E., G. Sakoulas, et al. (2003). "Comparison of mortality associated with methicillin-resistant and methicillin-susceptible Staphylococcus aureus bacteremia: a meta-analysis." Clin Infect Dis 36(1): 53-59.

Couto, I., S. W. Wu, et al. (2003). "Development of methicillin resistance in clinical isolates of Staphylococcus sciuri by transcriptional activation of the mecA homologue native to s." J Bacteriol 185(2): 645-653.

Crisostomo, M. I., H. Westh, et al. (2001). "The evolution of methicillin resistance in Staphylococcus aureus: similarity of genetic backgrounds in historically early methicillin-susceptible and -resistant isolates and contemporary epidemic clones." Proc Natl Acad Sci U S A 98(17): 9865-9870.

Croucher, N. J., S. R. Harris, et al. (2011). "Rapid pneumococcal evolution in response to clinical interventions." Science 331(6016): 430-434.

Dale, J. and S. Park (2004). Molecular genetics of bacteria. Chichester, John Wiley & Sons.

Darling, A. E., B. Mau, et al. (2010). "progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement." PLoS ONE 5(6): e11147.

Das, I., N. O'Connell, et al. (2007). "Epidemiology, clinical and laboratory characteristics of Staphylococcus aureus bacteraemia in a university hospital in UK." J Hosp Infect 65(2): 117-123.

David, M. D., A. M. Kearns, et al. (2006). "Community-associated meticillin-resistant Staphylococcus aureus: nosocomial transmission in a neonatal unit." J Hosp Infect 64(3): 244-250.

Davis, K. A., J. J. Stewart, et al. (2004). "Methicillin-resistant Staphylococcus aureus (MRSA) nares colonization at hospital admission and its effect on subsequent MRSA infection." Clin Infect Dis 39(6): 776-782.

Davis, S. L., M. B. Perri, et al. (2007). "Epidemiology and Outcomes of Community-Associated Methicillin-Resistant Staphylococcus aureus Infection." J Clin Microbiol.

Day, N. P., C. E. Moore, et al. (2002). "Retraction." Science 295(5557): 971.

Day, N. P., C. E. Moore, et al. (2001). "A link between virulence and ecological abundance in natural populations of Staphylococcus aureus." Science 292(5514): 114-116.

Deleo, F. R., M. Otto, et al. (2010). "Community-associated meticillin-resistant Staphylococcus aureus." Lancet 375(9725): 1557-1568.

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

Diederen, B. M. and J. A. Kluytmans (2006). "The emergence of infections with community-associated methicillin resistant Staphylococcus aureus." J Infect 52(3): 157-168.

Diekema, D. J., M. A. Pfaller, et al. (2001). "Survey of infections due to Staphylococcus species: frequency of occurrence and antimicrobial susceptibility of isolates collected in the United States, Canada, Latin America, Europe, and the Western Pacific region for the SENTRY Antimicrobial Surveillance Program, 1997-1999." Clin Infect Dis 32 Suppl 2: S114-132.

Diep, B. A., H. A. Carleton, et al. (2006). "Roles of 34 virulence genes in the evolution of hospital- and community-associated strains of methicillin-resistant Staphylococcus aureus." J Infect Dis 193(11): 1495-1503.

Diep, B. A., S. R. Gill, et al. (2006). "Complete genome sequence of USA300, an epidemic clone of community-acquired meticillin-resistant Staphylococcus aureus." Lancet 367(9512): 731-739.

Diep, B. A., A. M. Palazzolo-Ballance, et al. (2008). "Contribution of Panton-Valentine leukocidin in community-associated methicillin-resistant Staphylococcus aureus pathogenesis." PLoS ONE 3(9): e3198.

Drummond, A. J. and A. Rambaut (2007). "BEAST: Bayesian evolutionary analysis by sampling trees." BMC Evol Biol 7: 214.

Ellington, M. J., R. Hope, et al. (2010). "Decline of EMRSA-16 amongst methicillin-resistant Staphylococcus aureus causing bacteraemias in the UK between 2001 and 2007." J Antimicrob Chemother 65(3): 446-448.

Elston, J. W. and G. D. Barlow (2009). "Community-associated MRSA in the United Kingdom." J Infect 59(3): 149-155.

Emonts, M., A. G. Uitterlinden, et al. (2008). "Host polymorphisms in interleukin 4, complement factor H, and C-reactive protein associated with nasal carriage of Staphylococcus aureus and occurrence of boils." J Infect Dis 197(9): 1244-1253.

Enright, M. C., N. P. Day, et al. (2000). "Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of Staphylococcus aureus." J Clin Microbiol 38(3): 1008-1015.

Enright, M. C., D. A. Robinson, et al. (2002). "The evolutionary history of methicillin-resistant Staphylococcus aureus (MRSA)." Proc Natl Acad Sci U S A 99(11): 7687-7692.

Eriksen, N. H., F. Espersen, et al. (1994). "Evaluation of methods for the detection of nasal carriage of Staphylococcus aureus." Apmis 102(6): 407-412.

Eriksen, N. H., F. Espersen, et al. (1995). "Carriage of Staphylococcus aureus among 104 healthy persons during a 19-month period." Epidemiol Infect 115(1): 51-60.

Excoffier, L., G. Laval, et al. (2005). "Arlequin (version 3.0): an integrated software package for population genetics data analysis." Evol Bioinform Online 1: 47-50.

Feil, E. J., J. E. Cooper, et al. (2003). "How clonal is Staphylococcus aureus?" J Bacteriol 185(11): 3307-3316.

Feil, E. J., B. C. Li, et al. (2004). "eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data." J Bacteriol 186(5): 1518-1530.

Ferry, T., T. Perpoint, et al. (2005). "Virulence determinants in Staphylococcus aureus and their involvement in clinical syndromes." Curr Infect Dis Rep 7(6): 420-428.

Fleischmann, R. D., M. D. Adams, et al. (1995). "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd." Science 269(5223): 496-512.

Fraise, A. P., K. Mitchell, et al. (1997). "Methicillin-resistant Staphylococcus aureus (MRSA) in nursing homes in a major UK city: an anonymized point prevalence survey." Epidemiol Infect 118(1): 1-5.

Gardy, J. L., J. C. Johnston, et al. (2011). "Whole-genome sequencing and social-network analysis of a tuberculosis outbreak." N Engl J Med 364(8): 730-739.

Gill, S. R., D. E. Fouts, et al. (2005). "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." J Bacteriol 187(7): 2426-2438.

Goetz, A., K. Posey, et al. (1999). "Methicillin-resistant Staphylococcus aureus in the community: a hospital-based study." Infect Control Hosp Epidemiol 20(10): 689-691.

Gomez, M. I., M. O'Seaghdha, et al. (2006). "Staphylococcus aureus protein A activates TNFR1 signaling through conserved IgG binding domains." J Biol Chem 281(29): 20190-20196.

Graffunder, E. M. and R. A. Venezia (2002). "Risk factors associated with nosocomial methicillin-resistant Staphylococcus aureus (MRSA) infection including previous use of antimicrobials." J Antimicrob Chemother 49(6): 999-1005.

Graham, P. L., 3rd, S. X. Lin, et al. (2006). "A U.S. population-based survey of Staphylococcus aureus colonization." Ann Intern Med 144(5): 318-325.

Gray, R. R., A. J. Tatem, et al. (2011). "Testing spatiotemporal hypothesis of bacterial evolution using Methicillin-Resistant Staphylococcus aureus ST239 genome-wide data within a Bayesian framework." Mol Biol Evol.

Groom, A. V., D. H. Wolsey, et al. (2001). "Community-acquired methicillin-resistant Staphylococcus aureus in a rural American Indian community." Jama 286(10): 1201-1205.

Grundmann, H., D. M. Aanensen, et al. (2010). "Geographic distribution of Staphylococcus aureus causing invasive infections in Europe: a molecular-epidemiological analysis." PLoS Med 7(1): e1000215.

Grundmann, H., M. Aires-de-Sousa, et al. (2006). "Emergence and resurgence of meticillin-resistant Staphylococcus aureus as a public-health threat." Lancet 368(9538): 874-885.

Grundmann, H., S. Hori, et al. (2001). "Determining confidence intervals when measuring genetic diversity and the discriminatory abilities of typing methods for microorganisms." J Clin Microbiol 39(11): 4190-4192.

Haagen, I. A., H. C. Heezius, et al. (1990). "Adherence of peritonitis-causing staphylococci to human peritoneal mesothelial cell monolayers." J Infect Dis 161(2): 266-273.

Halablab, M. A., S. M. Hijazi, et al. (2010). "Staphylococcus aureus nasal carriage rate and associated risk factors in individuals in the community." Epidemiol Infect 138(5): 702-706.

Hallin, M., O. Denis, et al. (2007). "Genetic relatedness between methicillin-susceptible and methicillin-resistant Staphylococcus aureus: results of a national survey." J Antimicrob Chemother 59(3): 465-472.

Hanssen, A. M. and J. U. Ericson Sollid (2006). "SCCmec in staphylococci: genes on the move." FEMS Immunol Med Microbiol 46(1): 8-20.

Harbarth, S., N. Liassine, et al. (2000). "Risk factors for persistent carriage of methicillin-resistant Staphylococcus aureus." Clin Infect Dis 31(6): 1380-1385.

Harmsen, D., H. Claus, et al. (2003). "Typing of methicillin-resistant Staphylococcus aureus in a university hospital setting by using novel software for spa repeat determination and database management." J Clin Microbiol 41(12): 5442-5448.

Harris, S. R., E. J. Feil, et al. (2010). "Evolution of MRSA during hospital transmission and intercontinental spread." Science 327(5964): 469-474.

Health Protection Agency. (2005). "Community MRSA in England and Wales: definition through strain characterisation." The Communicable Disease Report Weekly 15(11): 3 - 4.

Health Protection Agency. (2007). "Antimicrobial Resistance in England, Wales and Northern Ireland, 2006."

Health Protection Agency. (2007). Commentary on Quarterly, Six-monthly and annual data for MRSA bacteraemia derived from mandatory surveillance, Health Protection Agency.

Health Protection Agency. (2007). "Identification of Staphylococcus species, micrococcus species and rothia species." National Standard Methods(2.1).

Health Protection Agency. (2009). Quarterly reports of MRSA bacteraemia (April 2001 - March 2007).

Herwaldt, L. A., J. J. Cullen, et al. (2004). "Preoperative risk factors for nasal carriage of Staphylococcus aureus." Infect Control Hosp Epidemiol 25(6): 481-484.

Hiramatsu, K., H. Hanaki, et al. (1997). "Methicillin-resistant Staphylococcus aureus clinical strain with reduced vancomycin susceptibility." J Antimicrob Chemother 40(1): 135-136.

Holden, M. T., E. J. Feil, et al. (2004). "Complete genomes of two clinical Staphylococcus aureus strains: evidence for the rapid evolution of virulence and drug resistance." Proc Natl Acad Sci U S A 101(26): 9786-9791.

Holmes, A., M. Ganner, et al. (2005). "Staphylococcus aureus isolates carrying Panton-Valentine leucocidin genes in England and Wales: frequency, characterization, and association with clinical disease." J Clin Microbiol 43(5): 2384-2390.

Huijsdens, X. W., B. J. van Dijke, et al. (2006). "Community-acquired MRSA and pig-farming." Ann Clin Microbiol Antimicrob 5: 26.

Huletsky, A., R. Giroux, et al. (2004). "New real-time PCR assay for rapid detection of methicillin-resistant Staphylococcus aureus directly from specimens containing a mixture of staphylococci." J Clin Microbiol 42(5): 1875-1884.

Hussain, F. M., S. Boyle-Vavra, et al. (2001). "Community-acquired methicillin-resistant Staphylococcus aureus colonization in healthy children attending an outpatient pediatric clinic." Pediatr Infect Dis J 20(8): 763-767.

Ito, T., X. X. Ma, et al. (2004). "Novel type V staphylococcal cassette chromosome mec driven by a novel cassette chromosome recombinase, ccrC." Antimicrob Agents Chemother 48(7): 2637-2651.

Iwase, T., Y. Uehara, et al. "Staphylococcus epidermidis Esp inhibits Staphylococcus aureus biofilm formation and nasal colonization." Nature 465(7296): 346-349.

Johnson, A. P., H. M. Aucken, et al. (2001). "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)." J Antimicrob Chemother 48(1): 143-144.

Katayama, Y., T. Ito, et al. (2000). "A new class of genetic element, staphylococcus cassette chromosome mec, encodes methicillin resistance in Staphylococcus aureus." Antimicrob Agents Chemother 44(6): 1549-1555.

Kennedy, A. D., M. Otto, et al. (2008). "Epidemic community-associated methicillin-resistant Staphylococcus aureus: Recent clonal expansion and diversification." Proc Natl Acad Sci U S A.

Kibbe, W. A. (2007). "OligoCalc: an online oligonucleotide properties calculator." Nucleic Acids Res 35(Web Server issue): W43-46.

Kirby, W. M. (1944). "Extraction of a Highly Potent Penicillin Inactivator from Penicillin Resistant Staphylococci." Science 99(2579): 452-453.

Kluytmans, J., A. van Belkum, et al. (1997). "Nasal carriage of Staphylococcus aureus: epidemiology, underlying mechanisms, and associated risks." Clin Microbiol Rev 10(3): 505-520.

Koreen, L., S. V. Ramaswamy, et al. (2004). "spa typing method for discriminating among Staphylococcus aureus isolates: implications for use of a single marker to detect genetic micro- and macrovariation." J Clin Microbiol 42(2): 792-799.

Kuehnert, M. J., D. Kruszon-Moran, et al. (2006). "Prevalence of Staphylococcus aureus nasal colonization in the United States, 2001-2002." J Infect Dis 193(2): 172-179.

Kuroda, M., T. Ohta, et al. (2001). "Whole genome sequencing of meticillin-resistant Staphylococcus aureus." Lancet 357(9264): 1225-1240.

Lan, R. and P. R. Reeves (2000). "Intraspecies variation in bacterial genomes: the need for a species genome concept." Trends Microbiol 8(9): 396-401.

Laurent, F., H. Lelievre, et al. (2001). "Fitness and competitive growth advantage of new gentamicin-susceptible MRSA clones spreading in French hospitals." J Antimicrob Chemother 47(3): 277-283.

Lee, S. M., M. Ender, et al. (2007). "Fitness cost of staphylococcal cassette chromosome mec in methicillin-resistant Staphylococcus aureus by way of continuous culture." Antimicrob Agents Chemother 51(4): 1497-1499.

Li, H., J. Ruan, et al. (2008). "Mapping short DNA sequencing reads and calling variants using mapping quality scores." Genome Res 18(11): 1851-1858.

Limbago, B., G. E. Fosheim, et al. (2009). "Characterization of methicillin-resistant Staphylococcus aureus isolates collected in 2005 and 2006 from patients with invasive disease: a population-based analysis." J Clin Microbiol 47(5): 1344-1351.

Lindsay, J. A. and M. T. Holden (2006). "Understanding the rise of the superbug: investigation of the evolution and genomic variation of Staphylococcus aureus." Funct Integr Genomics 6(3): 186-201.

Lindsay, J. A., C. E. Moore, et al. (2006). "Microarrays reveal that each of the ten dominant lineages of Staphylococcus aureus has a unique combination of surface-associated and regulatory genes." J Bacteriol 188(2): 669-676.

Lodise, T. P., Jr., P. S. McKinnon, et al. (2003). "Prediction model to identify patients with Staphylococcus aureus bacteremia at risk for methicillin resistance." Infect Control Hosp Epidemiol 24(9): 655-661.

Lowy, F. D. (1998). "Staphylococcus aureus infections." N Engl J Med 339(8): 520-532.

Lowy, F. D. (2003). "Antimicrobial resistance: the example of Staphylococcus aureus." J Clin Invest 111(9): 1265-1273.

Lucet, J. C., X. Paoletti, et al. (2009). "Carriage of methicillin-resistant Staphylococcus aureus in home care settings: prevalence, duration, and transmission to household members." Arch Intern Med 169(15): 1372-1378.

Lunter, G. and M. Goodson (2010). "Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads." Genome Res.

MacKinnon, M. M. and K. D. Allen (2000). "Long-term MRSA carriage in hospital patients." J Hosp Infect 46(3): 216-221.

Maiden, M. C. (2006). "Multilocus Sequence Typing of Bacteria." Annu Rev Microbiol.

Maiden, M. C., J. A. Bygraves, et al. (1998). "Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms." Proc Natl Acad Sci U S A 95(6): 3140-3145.

Malachowa, N. and F. R. DeLeo (2010). "Mobile genetic elements of Staphylococcus aureus." Cell Mol Life Sci 67(18): 3057-3071.

Malachowa, N., A. Sabat, et al. (2005). "Comparison of multiple-locus variable-number tandem-repeat analysis with pulsed-field gel electrophoresis, spa typing, and multilocus sequence typing for clonal characterization of Staphylococcus aureus isolates." J Clin Microbiol 43(7): 3095-3100.

Mandell, G. L. B., John E; Raphael, Dolin (2005). Mandell, Douglas and Bennett's Principles and Practice of Infectious Diseases.

Manzur, A., M. Vidal, et al. (2007). "Predictive factors of meticillin resistance among patients with Staphylococcus aureus bloodstream infections at hospital admission." J Hosp Infect 66(2): 135-141.

Marschall, J. and K. Muhlemann (2006). "Duration of methicillin-resistant Staphylococcus aureus carriage, according to risk factors for acquisition." Infect Control Hosp Epidemiol 27(11): 1206-1212.

McClure, J. A., J. M. Conly, et al. (2006). "Novel multiplex PCR assay for detection of the staphylococcal virulence marker Panton-Valentine leukocidin genes and simultaneous discrimination of methicillin-susceptible from -resistant staphylococci." J Clin Microbiol 44(3): 1141-1144.

McFarlan, A. M. (1938). "Pathogenic Staphylococci in the Nose." Br Med J 2(4061): 939-941.

Medini, D., C. Donati, et al. (2005). "The microbial pan-genome." Curr Opin Genet Dev 15(6): 589-594.

Melles, D. C., R. F. Gorkink, et al. (2004). "Natural population dynamics and expansion of pathogenic clones of Staphylococcus aureus." J Clin Invest 114(12): 1732-1740.

Mellmann, A., T. Weniger, et al. (2007). "Based Upon Repeat Pattern (BURP): an algorithm to characterize the long-term evolution of Staphylococcus aureus populations based on spa polymorphisms." BMC Microbiol 7: 98.

Millar, B. C., A. Loughrey, et al. (2007). "Proposed definitions of community-associated meticillin-resistant Staphylococcus aureus (CA-MRSA)." J Hosp Infect 67(2): 109-113.

Miller, M., H. A. Cook, et al. (2009). "Staphylococcus aureus in the community: colonization versus infection." PLoS One 4(8): e6708.

Moran, G. J., A. Krishnadasan, et al. (2006). "Methicillin-resistant S. aureus infections among patients in the emergency department." N Engl J Med 355(7): 666-674.

Munckhof, W. J., G. R. Nimmo, et al. (2009). "Nasal carriage of Staphylococcus aureus, including community-associated methicillin-resistant strains, in Queensland adults." Clin Microbiol Infect 15(2): 149-155.

Naimi, T. S., K. H. LeDell, et al. (2003). "Comparison of community- and health care-associated methicillin-resistant Staphylococcus aureus infection." Jama 290(22): 2976-2984.

Nashev, D., K. Toshkova, et al. (2007). "Distribution of enterotoxin genes among carriage- and infection-associated isolates of Staphylococcus aureus." Lett Appl Microbiol 45(6): 681-685.

Nature (2010). "Human genome at ten: The sequence explosion." Nature 464(7289): 670-671.

Noble, W. C., H. A. Valkenburg, et al. (1967). "Carriage of Staphylococcus aureus in random samples of a normal population." J Hyg (Lond) 65(4): 567-573.

Noto, M. J., B. N. Kreiswirth, et al. (2007). "Gene Acquisition at the Insertion Site for SCCmec, the Genomic Island Conferring Methicillin Resistance in Staphylococcus aureus." J Bacteriol.

Nouwen, J. L., A. Ott, et al. (2004). "Predicting the Staphylococcus aureus nasal carrier state: derivation and validation of a "culture rule"." Clin Infect Dis 39(6): 806-811.

Novick, R. P. (2003). "Mobile genetic elements and bacterial toxinoses: the superantigen-encoding pathogenicity islands of Staphylococcus aureus." Plasmid 49(2): 93-105.

Nubel, U., P. Roumagnac, et al. (2008). "Frequent emergence and limited geographic dispersal of methicillin-resistant Staphylococcus aureus." Proc Natl Acad Sci U S A 105(37): 14130-14135.

Nulens, E., I. Gould, et al. (2005). "Staphylococcus aureus carriage among participants at the 13th European Congress of Clinical Microbiology and Infectious Diseases." Eur J Clin Microbiol Infect Dis 24(2): 145-148.

O'Sullivan, N. P. and C. T. Keane (2000). "Risk factors for colonization with methicillin-resistant Staphylococcus aureus among nursing home residents." J Hosp Infect 45(3): 206-210.

O'Sullivan, N. R. and C. T. Keane (2000). "The prevalence of methicillin-resistant staphylococcus aureus among the residents of six nursing homes for the elderly." J Hosp Infect 45(4): 322-329.

Office for National Statistics. (2007). "Population and vital statistics by area of usual residence in the United Kingdom, 2005." Key Population and Vital Statistics 32(PP1 No 28).

Ogston, A. (1984). "Classics in infectious diseases. "On abscesses". Alexander Ogston (1844-1929)." Rev Infect Dis 6(1): 122-128.

Okuma, K., K. Iwakawa, et al. (2002). "Dissemination of new methicillin-resistant Staphylococcus aureus clones in the community." J Clin Microbiol 40(11): 4289-4294.

Oliveira, D. C., A. Tomasz, et al. (2002). "Secrets of success of a human pathogen: molecular evolution of pandemic clones of meticillin-resistant Staphylococcus aureus." Lancet Infect Dis 2(3): 180-189.

Olsen, J. E., H. Christensen, et al. (2006). "Diversity and evolution of blaZ from Staphylococcus aureus and coagulase-negative staphylococci." J Antimicrob Chemother 57(3): 450-460.

Orendi, J. (2006). "Hospital-associated transmission of Panton-Valentine leukocidin (PVL) positive community-associated MRSA in the West Midlands." Communicable Disease Report Weekly 16(50): 2-3.

Otter, J. A. and G. L. French (2008). "The emergence of community-associated methicillin-resistant Staphylococcus aureus at a London teaching hospital, 2000-2006." Clin Microbiol Infect 14(7): 670-676.

Oztoprak, N., M. A. Cevik, et al. (2006). "Risk factors for ICU-acquired methicillin-resistant Staphylococcus aureus infections." Am J Infect Control 34(1): 1-5.

Page, R. D. M. and E. C. Holmes (1998). Molecular evolution : a phylogenetic approach. Oxford, Blackwell Science.

Pan, E. S., B. A. Diep, et al. (2005). "Population dynamics of nasal strains of methicillin-resistant Staphylococcus aureus--and their relation to community-associated disease activity." J Infect Dis 192(5): 811-818.

Peacock, S. J., A. Justice, et al. (2003). "Determinants of acquisition and carriage of Staphylococcus aureus in infancy." J Clin Microbiol 41(12): 5718-5725.

Peacock, S. J., C. E. Moore, et al. (2002). "Virulent combinations of adhesin and toxin genes in natural populations of Staphylococcus aureus." Infect Immun 70(9): 4987-4996.

Popovich, K. J., R. A. Weinstein, et al. (2008). "Are community-associated methicillin-resistant Staphylococcus aureus (MRSA) strains replacing traditional nosocomial MRSA strains?" Clin Infect Dis 46(6): 787-794.

Price, J., G. Baker, et al. (2010). "Clinical and Microbiological Determinants of Outcome in Staphylococcus aureus Bacteraemia." Int J Microbiol 2010: 654858.

Pritchard, J. K., M. Stephens, et al. (2000). "Inference of population structure using multilocus genotype data." Genetics 155(2): 945-959.

Projan, S. J. N., Richard P (1997). The Molecular Basis of Pathogenicity. The Staphylococci in Human Disease. K. B. A. Crossley, Gordon L, Churchill Livingstone Inc: 55-81.

Qu, F., E. Cui, et al. (2010). "Nasal colonization of and clonal transmission of methicillin-susceptible Staphylococcus aureus among Chinese military volunteers." J Clin Microbiol 48(1): 64-69.

Rezende, N. A., H. M. Blumberg, et al. (2002). "Risk factors for methicillin-resistance among patients with Staphylococcus aureus bacteremia at the time of hospital admission." Am J Med Sci 323(3): 117-123.

Robinson, D. A. and M. C. Enright (2003). "Evolutionary models of the emergence of methicillin-resistant Staphylococcus aureus." Antimicrob Agents Chemother 47(12): 3926-3934.

Robinson, D. A. and M. C. Enright (2004). "Evolution of Staphylococcus aureus by large chromosomal replacements." J Bacteriol 186(4): 1060-1064.

Robinson, D. A. and M. C. Enright (2004). "Multilocus sequence typing and the evolution of methicillin-resistant Staphylococcus aureus." Clin Microbiol Infect 10(2): 92-97.

Roghmann, M. C. (2000). "Predicting methicillin resistance and the effect of inadequate empiric therapy on survival in patients with Staphylococcus aureus bacteremia." Arch Intern Med 160(7): 1001-1004.

Safdar, N. and E. A. Bradley (2008). "The risk of infection after nasal colonization with Staphylococcus aureus." Am J Med 121(4): 310-315.

Sakwinska, O., D. S. Blanc, et al. (2010). "Ecological temporal stability of Staphylococcus aureus nasal carriage." J Clin Microbiol 48(8): 2724-2728.

Salgado, C. D., B. M. Farr, et al. (2003). "Community-acquired methicillin-resistant Staphylococcus aureus: a meta-analysis of prevalence and risk factors." Clin Infect Dis 36(2): 131-139.

Sanford, M. D., A. F. Widmer, et al. (1994). "Efficient detection and long-term persistence of the carriage of methicillin-resistant Staphylococcus aureus." Clin Infect Dis 19(6): 1123-1128.

Sanger, F., S. Nicklen, et al. (1977). "DNA sequencing with chain-terminating inhibitors." Proc Natl Acad Sci U S A 74(12): 5463-5467.

Sauerbrei, P. R. a. W. (2008). MFP: Multivariable Model-Building with Fractional Polynomials. Multivariable Modelling: A pragmatic approach based on fractional polynomials for continuous variables, Wiley.

Scanvic, A., L. Denic, et al. (2001). "Duration of colonization by methicillin-resistant Staphylococcus aureus after hospital discharge and risk factors for prolonged carriage." Clin Infect Dis 32(10): 1393-1398.

Schito, G. C. (2006). "The importance of the development of antibiotic resistance in Staphylococcus aureus." Clin Microbiol Infect 12 Suppl 1: 3-8.

Schmitz, F. J., M. E. Jones, et al. (1998). "Characterization of grlA, grlB, gyrA, and gyrB mutations in 116 unrelated isolates of Staphylococcus aureus and effects of mutations on ciprofloxacin MIC." Antimicrob Agents Chemother 42(5): 1249-1252.

Shendure, J. and H. Ji (2008). "Next-generation DNA sequencing." Nat Biotechnol 26(10): 1135-1145.

Sievert, D., Boulton, ML, Stoltman, G, Johnson, D, Stobierski, MG, Downes, FP, Somsel, PA, Rudrik, JT, Brown, W, Hafeez, W, Lundstrom, T, Flanagan, E, Johnson, R, Mitchell, J, Chang, S (2002). "Staphylococcus aureus resistant to vancomycin--United States, 2002." MMWR Morb Mortal Wkly Rep 51(26): 565-567.

Sivaraman, K., N. Venkataraman, et al. (2009). "Staphylococcus aureus nasal carriage and its contributing factors." Future Microbiol 4: 999-1008.

Spratt, B. G., W. P. Hanage, et al. (2004). "Displaying the relatedness among isolates of bacterial species -- the eBURST approach." FEMS Microbiol Lett 241(2): 129-134.

Srivatsan, A., Y. Han, et al. (2008). "High-precision, whole-genome sequencing of laboratory strains facilitates genetic studies." PLoS Genet 4(8): e1000139.

Staden, R. (1996). "The Staden sequence analysis package." Mol Biotechnol 5(3): 233-241.

Strohl, W. A. R., Harriet; Fisher, Bruce (2001). Lippincott's Illustrated Reviews: Microbiology, Lippincott, Williams, Wilkins.

Strommenger, B., C. Kettlitz, et al. (2006). "Assignment of Staphylococcus isolates to groups by spa typing, SmaI macrorestriction analysis, and multilocus sequence typing." J Clin Microbiol 44(7): 2533-2540.

Tacconelli, E., L. Venkataraman, et al. (2004). "Methicillin-resistant Staphylococcus aureus bacteraemia diagnosed at hospital admission: distinguishing between community-acquired versus healthcare-associated strains." J Antimicrob Chemother 53(3): 474-479.

Takano, T., W. Higuchi, et al. (2007). "Novel Characteristics of Community-Acquired Methicillin-Resistant Staphylococcus aureus Belonging to Multilocus Sequence Type 59 in Taiwan." Antimicrob Agents Chemother.

Tenover, F. C., R. D. Arbeit, et al. (1995). "Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing." J Clin Microbiol 33(9): 2233-2239.

Thomas, S., J. A. Karas, et al. (2007). "Meticillin-resistant Staphylococcus aureus carriage among district nurse patients and medical admissions in a UK district." J Hosp Infect 66(4): 369-373.

Tiemersma, E. W., S. L. Bronzwaer, et al. (2004). "Methicillin-resistant Staphylococcus aureus in Europe, 1999-2002." Emerg Infect Dis 10(9): 1627-1634.

Travers, K. J., C. S. Chin, et al. (2010). "A flexible and efficient template format for circular consensus sequencing and SNP detection." Nucleic Acids Res 38(15): e159.

Trindade, P. A., J. A. McCulloch, et al. (2003). "Molecular techniques for MRSA typing: current issues and perspectives." Braz J Infect Dis 7(1): 32-43.

Tristan, A., T. Ferry, et al. (2007). "Virulence determinants in community and hospital meticillin-resistant Staphylococcus aureus." Journal of Hospital Infection 65: 105-109.

van Belkum, A., D. C. Melles, et al. (2006). "Clonal distribution and differential occurrence of the enterotoxin gene cluster, egc, in carriage- versus bacteremia-associated isolates of Staphylococcus aureus." J Clin Microbiol 44(4): 1555-1557.

Van Belkum, A., N. H. Riewarts Eriksen, et al. (1997). "Coagulase and protein A polymorphisms do not contribute to persistence of nasal colonisation by Staphylococcus aureus." J Med Microbiol 46(3): 222-232.

van Belkum, A., N. J. Verkaik, et al. (2009). "Reclassification of Staphylococcus aureus nasal carriage types." J Infect Dis 199(12): 1820-1826.

van den Akker, E. L., J. L. Nouwen, et al. (2006). "Staphylococcus aureus nasal carriage is associated with glucocorticoid receptor gene polymorphisms." J Infect Dis 194(6): 814-818.

Van den Bergh, M. F., E. P. Yzerman, et al. (1999). "Follow-up of Staphylococcus aureus nasal carriage after 8 years: redefining the persistent carrier state." J Clin Microbiol 37(10): 3133-3140.

Van Domselaar, G. H., P. Stothard, et al. (2005). "BASys: a web server for automated bacterial genome annotation." Nucleic Acids Res 33(Web Server issue): W455-459.

Van den Bergh, M. F., E. P. Yzerman, et al. (1999). "Follow-up of Staphylococcus aureus nasal carriage after 8 years: redefining the persistent carrier state." J Clin Microbiol 37(10): 3133-3140.

Vandenesch, F., T. Naimi, et al. (2003). "Community-acquired methicillin-resistant Staphylococcus aureus carrying Panton-Valentine leukocidin genes: worldwide emergence." Emerg Infect Dis 9(8): 978-984.

Verena Schneider-Lindner, J. A. D., Sandra Dial, Andre Dascal, and Samy Suissa (2007). "Antimicrobial Drugs and Community-acquired Methicillin-Resistant Staphylococcus aureus, United Kingdom." Emerging Infectious Diseases 13(7): 994 - 1000.

von Eiff, C., K. Becker, et al. (2001). "Nasal carriage as a source of Staphylococcus aureus bacteremia." New England Journal of Medicine 344(1): 11-16.

Vos, M. and X. Didelot (2009). "A comparison of homologous recombination rates in bacteria and archaea." ISME J 3(2): 199-208.

Voyich, J. M., M. Otto, et al. (2006). "Is Panton-Valentine leukocidin the major virulence determinant in community-associated methicillin-resistant Staphylococcus aureus disease?" J Infect Dis 194(12): 1761-1770.

Vriens, M. R., H. E. Blok, et al. (2005). "Methicillin-resistant Staphylococcus aureus carriage among patients after hospital discharge." Infect Control Hosp Epidemiol 26(7): 629-633.

Washer, P. and H. Joffe (2006). "The "hospital superbug": social representations of MRSA." Soc Sci Med 63(8): 2141-2152.

Watterson, G. A. (1975). "On the number of segregating sites in genetical models without recombination." Theor Popul Biol 7(2): 256-276.

Wertheim, H. F., W. B. van Leeuwen, et al. (2005). "Associations between Staphylococcus aureus Genotype, Infection, and In-Hospital Mortality: A Nested Case-Control Study." J Infect Dis 192(7): 1196-1200.

Wertheim, H. F., M. C. Vos, et al. (2004). "Risk and outcome of nosocomial Staphylococcus aureus bacteraemia in nasal carriers versus non-carriers." Lancet 364(9435): 703-705.

Williams, R. E. (1963). "Healthy carriage of Staphylococcus aureus: its prevalence and importance." Bacteriol Rev 27: 56-71.

Witte, W., C. Cuny, et al. (2008). "Emergence and spread of antibiotic-resistant Gram-positive bacterial pathogens." Int J Med Microbiol 298(5-6): 365-377.

Wyllie, D. H., D. W. Crook, et al. (2006). "Mortality after Staphylococcus aureus bacteraemia in two hospitals in Oxfordshire, 1997-2003: cohort study." Bmj 333(7562): 281.

Wyllie, D. H., A. S. Walker, et al. (2007). "Hospital exposure in a UK population, and its association with bacteraemia." J Hosp Infect 67(4): 301-307.

Yang, Z. and J. P. Bielawski (2000). "Statistical methods for detecting molecular adaptation." Trends Ecol Evol 15(12): 496-503.

Zerbino, D. R. and E. Birney (2008). "Velvet: algorithms for de novo short read assembly using de Bruijn graphs." Genome Res 18(5): 821-829.

Zhang, K., J. A. McClure, et al. (2009). "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." Antimicrob Agents Chemother 53(2): 531-540.

8 Appendices

Appendix 1: Protocol for *S. aureus* isolation from a nasal swab

1. If the swab is the first swab:
    a. Label 5% NaCl broths with matching consecutive numbers to the swab, all ending in 0
    b. Label the MySwabID column in the swab spreadsheet with the same numbers as the NaCl broths
    c. Order swabs in same orientation as 5% NaCl broths
    d. Clean scissors used to cut swab with 70% ethanol
    e. Open swab vial and cut swab into 5% NaCl broth then close lid
2. If the swab is a follow-up swab
    a. Label 5% NaCl broths with numbers on return envelopes
    b. Open each envelope in turn and check number on follow-up sheet matches that on envelope
    c. Write corresponding number in MySwabID column on swab spreadsheet
    d. Clean scissors used to cut swab with 70% ethanol
    e. Open swab pack and swab vial
    f. Cut swab into 5% NaCl broth then close lid
3. Dispose of all swab packaging in lab bin
4. Place swab and 5% NaCl broth at 37 ^{0}C overnight
5. Clean bench with Virkon

---

6. Label SaSelect agar with same numbers from swabs put in broth the previous day (copy this from swab spreadsheet not broth bottle to avoid error)
7. Sub-culture a loopful of broth onto SaSelect agar with corresponding number
   Pattern for culture onto the agar:

8. Place SaSelect agar at 37 ^{0}C overnight

---

9. Record SaSelect agars with growth of pink/orange colonies as PINK in SaSelect column of swab spreadsheet. Record SaSelect agars with no growth of pink/orange colonies as NG in SaSelect column on swab spreadsheet
10. If there are only a few pink colonies present or there is mixed growth with other coloured colonies making it hard to isolate a single pink colony then the following should be performed:
    a. Label a new SaSelect agar with the corresponding number

b. Using a loop, touch as many pink colonies as possible from the original SaSelect agar
c. Emulsify and dilute the growth on the loop in peptone water
d. Take a loopful of the peptone water and sub-culture onto the newly labelled SaSelect agar
e. Incubate at 37 ^{0}C overnight

11. Discard SaSelect agar without growth of pink/orange colonies, take those with pink/orange colonies for further testing

---

Further testing 1:

12. Divide each Columbia blood agar with 5% NaCl (CBS) into three sections and each DNase agar into eight parts
13. Label one positive and negative control on each DNase agar (+ve = Oxford Staph positive control, -ve = DNAse negative)
14. Label one MSSA control on one third of a CBS
15. Using the swab spreadsheet as a reference for each "PINK" swab, label
    a. 1 CBA
    b. 1 third of a CBS
    c. 1 eighth of a DNase agar
16. Emulsify 1 pink/orange colony from SaSelect agar into peptone water
17. Use a swab to culture this onto one third of a CBS
18. Using forceps place an oxacillin disk in the middle of the section
19. Take a loopfull of the peptone water solution and culture onto CBA to confirm purity

20. Take another loopfull of the peptone water solution and culture a line onto an eighth of the DNAse agar
21. Mark the peptone water bottle with a cross to ensure it does not get re-used
22. Leave completed CBS, CBA and DNase at 37 ^{0}C overnight

Further testing 2:

23. If growth on the CBA is not pure, take one colony from the CBA and emulsify into peptone water. Then repeat from step 15

    If CBA pure:

24. Tick the box in the Pure from Peptone column on the swab spreadsheet
25. Catalase test on colonies from CBA
    a. Using a capillary take a small volume of hydrogen peroxide into the tube
    b. Place the tube onto a *S. aureus* colony on the CBA so a little *S. aureus* goes into the tube
    c. If bubbles pass up the tube the bacteria is catalase positive, tick the Catalase column on the swab spreadsheet

d. If there are no bubbles the bacteria is catalase negative and therefore not *S. aureus,* write –ve in the Catalase column in the swab spreadsheet

26. Complete DNase test
    a. Pour enough 1M HCl onto the DNase agar (prepared the previous day) to cover the agar
    b. Leave for about 10 min
    c. If there is an area of clearing around the bacteria it is DNase positive, tick the DNase column in the *S. aureus* spreadsheet
    d. If there is no area of clearing around the bacteria it is not *S. aureus*, write –ve in the DNAse column in the *S. aureus* spreadsheet
    e. Use the positive and negative control to confirm the agar is working and compare the clear areas
27. Staphaurex test on colonies from CBA
    a. Using the card provided with the Staphaurex kit put a small dot of the negative solution (blue) on the left circle and a small dot of the positive solution (red) on the right circle
    b. Using a wooden stick put a little of the bacteria from the CBA onto the negative solution and emulsify into the solution for about 10 s
    c. Similarly use the same wooden stick to put another small amount of the bacteria from the CBA into the positive solution
    d. If the positive solution agglutinates and the negative solution does not the bacteria is Staphaurex positive, tick the Staphaurex column in the *S. aureus* spreadsheet
    e. If the positive solution does not agglutinate the bacteria is not *S. aureus,* write –ve in the Staphaurex column on the *S. aureus* spreadsheet
    f. If the positive solution does not agglutinate (but sample is catalase and DNAase positive*)* carry out a tube coagulase test (see point 28)
28. Tube coagulase
    a. Take three bottles of peptone
    b. Label these positive, negative and appropriate swab number
    c. Add 0.5 ml rabbit plasma (from 4 ^{0}C) to each bottle
    d. Take a colony of Oxford Staph control, and emulsify into the tube labelled positive
    e. Take a colony from the CBA and emulsify into the bottle with the swab number
    f. Incubate all three bottles at 37 ^{0}C for at least four hours
    g. After four hours look at tubes, and see if the peptone water has agglutinated (as the positive control should have)
29. If the strain has ticks in the Catalase, Staphaurex, DNAse and Pure from Peptone columns on the *S. aureus* spreadsheet it is *S. aureus*
30. Determine whether the *S. aureus* is meticillin resistant or sensitive using the CBS
    a. If there is an area of no growth around the oxacillin disk on the CBS $\geq$ 14mm in diameter the *S. aureus* is MSSA, put SS in the Oxacillin column on the swab spreadsheet
    b. If there is no area of no growth around the oxacillin disk on the CBS of the area of no growth has a diameter $\leq$ 14mm the *S. aureus* is MRSA, put R in the Oxacillin column on the swab spreadsheet

c. If the area of no growth is too similar to 14mm to determine an oxacillin Etest must be performed to confirm whether the *S. aureus* is MSSA or MRSA. Once the sensitivity is confirmed, mark the swab spreadsheet as appropriate.

31. If the *S. aureus* is MRSA for the first time
    a. Re-culture one colony from the CBA onto a new CBA
    b. Leave at 37 ^{0}C overnight
    c. Take a charcoal swab and rub over the CBA to pick up as much of the presumed MRSA as possible
    d. Take charcoal swab to routine lab for confirmation of MRSA
32. If confirmed *S. aureus* (MRSA or MSSA)
33. Make a glycerol bacteria stock and DNA boilate
    a. For the *S. aureus* positive strains on the swab spreadsheet label the lids of
        i. 4 x 200 µl PCR tubes
        ii. 1 x 1.5 ml epindorf
        iii. 3 x Tubee's high low strips

    with the name of the sample
    b. Stick the Tubee's strip onto the appropriate PCR tube
    c. Add 50 µl 1 x TE buffer to the 200 µl PCR tube without the Tubee label
    d. Add 400 µl Saline to the 1. 5ml epindorf
    e. Using a 5 µl loop, emulsify a loopful of pink colonies (approximately 50 different colonies) from previous SaSelect agar into the saline in the correct epindorf
    f. Remove 10 µl of the emulsified *S. aureus* saline solution and add to the 0.2 µl PCR tube containing 5 0µl 1 x TE buffer and place to one side (see step 34 for the next stage in making a DNA boilate)
    g. Add 200 µl 45% glycerol to the 1.5 ml epindorf containing the emulsified *S. aureus* saline solution
    h. Vortex the epindorf
    i. From the epindorf aliquot 200 µl of the solution into 2 x 200 µl PCR tubes (one master one and working stock)
    j. Take each pair of PCR tubes and put one in a working stock box and one in a master stock box in the -80 ^{0}C freezer
    k. After making the stock tick the MIXED stock made column on the swab spreadsheet and write which storage box and location within the box that the stock is stored
    l. Store the boxes at 80 ^{0}C
34. Make a DNA boilate (continued)
    a. Place the PCR tube which was set aside earlier in the PCR machine at 99.9 ^{0}C for 10 minutes
    b. Centrifuge at 13.2 rpm for 2 min
    c. Take 50 µl supernatant from the PCR tube and transfer into the remaining empty 200 µl PCR tube with the same label
    d. After making the boilate tick the MIXED DNA boilate made box on the swab spreadsheet
    e. Take the PCR tube and store in a box at -20 ^{0}C. The box and position should be the same as the bacterial stock

Appendix 2: Admission Bacteraemia questionnaire

Case number MRSA☐ SAUR☐

Hospital No.
Name (Surname/Forname)

Date of birth Male☐ Female☐

Admission:

Admission date:
Admitted from: Nursing home ☐ Residential Home☐
Own home ☐ Other☐

Has the patient been admitted to the ORH before?
Y☐ → Date of Discharge
N☐ From which department?
Unknown☐

No. of Outpatient appts in last year at ORH . Last OPD

No. of Inpatient appts in last year at ORH:

Has the patient been admitted to another hospital? Y☐ When No☐ Unknown☐

Has the patient ever had: Renal dialysis☐ Chemotherapy☐ ITU admission☐

Comorbidities on admission

Diabetes Mellitus☐	Heart failure☐
Current Oral steroids☐	Vascular disease☐
COPD☐	Chronic renal failure☐

Has the patient had surgery in the last year? Y☐ N☐ Unknown☐
Has the patient had invasive intervention in the last year? Y☐ What? N☐
Unknown☐

On admission, did the patient have a record of currently having…

Vascular access (central line, Hickman line, PICC line)	Y☐	N☐	Unknown☐
Urinary Catheter (recorded in nursing notes)	Y☐	N☐	Unknown☐
Surface ulcers of other skin breaks	Y☐	N☐	Unknown☐
Intravenuous Drug Use	Y☐	N☐	Unknown☐
Eczema/Psoriasis	Y☐	N☐	Unknown☐

MRSA status on PAS Prev. isolated☐ Not prev. isolated☐
Blood culture result known at time of admission Y☐ N☐

Presenting complaint
Presenting speciality
Referred from GP☐ A+E☐ Inter-hospital transfer☐
Direct admission☐ Why

Was infection suspected in admission clerking? Y☐ Source? N☐

Admission presentation (from first medical clerking or chart)

T (C) BP Pulse

Admission bloods

White cell count x 10^9/L CRP

Neutrophils x 10^9/L Platelets x 10^9/L

Lymphocytes x 10^9/L Haemoglobin g/dl

MEDICATION: New antibiotics started in the ***first 4 days*** since admission

1. Drug name Route Date started Date stopped (if in first 4 days)
2. Drug name Route Date started Date stopped (if in first 4 days)
3. Drug name Route Date started Date stopped (if in first 4 days)
4. Drug name Route Date started Date stopped (if in first 4 days)
5. Drug name Route Date started Date stopped (if in first 4 days)

Any documented antibiotic allergies? Y☐ What? N☐ Unknown☐

OUTCOME:

ITU Admission? Y☐ Why? N☐ Unknown☐

Death? Y☐ Date N☐ Unknown☐

Case carried out by

Extra information

Appendix 3: GP questionnaire

NHS	Department of Microbiology John Radcliffe Hospital Oxford OX3 9DU FAX number: 01865 764192 Telephone Number:01865 220879 / 220884

Dr
Surgery Name
Surgery address1
Surgery address2

Date:

Dear Dr xxxx,

SURNAME, First, DOB: 00/00/1900, Hosp No.: 0000000
Specimen reference: xxxx, Specimen date: 00/00/2007

You will have already received a formal report of methicillin-resistant *Staphylococcus aureus* from a specimen from the above patient.

We are carrying out an audit of patients who fit the criteria of having potential community acquired MRSA. Our records show that there was no acute hospital contact for this patient in the 12 months prior to this date.

Please find attached a questionnaire, which I would be very grateful if you could complete. We are interested in information relating to the 12 months before the patient had MRSA isolated, i.e. between xx xxxxxxx 2006 to xx xxxxxxxxxx 2007.

If you would like to discuss the clinical management of this patient, please contact Dr Esmail on 01865 741166 bleep 4055.

Thank you for your help.

Yours sincerely

Ruth Miller

Department of Microbiology
John Radcliffe Hospital
Oxford OX3 9DU
FAX number: 01865 764192
Telephone Number:01865 220879 / 220884

SURNAME, First, DOB: 00/00/1900, Hosp No.: 0000000
Specimen reference: xxxx, Specimen date: 00/00/2007

1. Has the patient been a hospital in-patient in the past 12 months? yes / no
   if yes, what hospital(s) and why?.....
   if yes, are they still an inpatient? yes / no

2. Has the patient been a hospital out-patient in the past 12 months? yes / no
   if yes, what hospital(s) and why?.....

3. If the patient was not in hospital in the past 12-months (questions 1 & 2 answered "no")
   when was the patient last in hospital?.....
   where was the patient last in hospital?.....

4. Has the patient been resident in a nursing or care home in the past 12 months? yes / no
   if yes, what type of residential home?.....
   if yes, are they still residing there? yes / no

5. Is the patient receiving regular care (e.g. wound dressing) from the practice nurse? yes / no
   if yes, what type of care?.....

6. Does the patient have an indwelling catheter or have they undergone any invasive procedures in the community in the past 12 months? yes / no
   if yes, what?. .....

7. Has the patient received antibiotics in the community in the past year? yes / no
   if yes, how long ago?.....

8. Has the patient had MRSA previously isolated? yes / no
   if yes, when was this?.....

If there is anything else you can add about this patient that you think has put them at risk of acquiring MRSA (e.g. known close contact with or at risk of MRSA) please give brief details below:

.....
.....
.....
.....

Appendix 4: Carriage study participant information sheet

Department of Primary Health Care

Oxford Radcliffe Hospitals NHS
NHS Trust

NDM Laboratory, Room 7724, Level 7, ORH Trust, Oxford, OX3 9DU
Tel: +44 (0)1865 222884, Fax: +44 (0)1865 222191
E-mail: staph@ndm.ox.ac.uk
Website: www.medsci.ox.ac.uk/staph

Study title: Population structure and carriage length of *Staphylococcus aureus*
Ethics committee number: 08/H0605/102
Chief Investigator: Dr Derrick Crook

Participant Information Sheet

1. Invitation

You are being invited to take part in a research study. It is important for you to understand why the research is being done and what the study will involve. Please take time to read the following information carefully and discuss it with others if you wish.

Part 1 of this sheet tells you the purpose of this study and what will happen to you if you take part.
Part 2 of this sheet gives you more detailed information about the conduct of the study.
Please feel free to contact the research team, whose telephone number is at the top of this letter, if there is anything that is not clear or if you would like more information. Thank you for considering taking part in this study.

Part 1

2. What is the purpose of the study?

a. The objective of the study is to follow-up people **to identify those** who carry the bacteria *Staphylococcus aureus* in their nose or elsewhere in their body.
b. In many people *Staphylococcus aureus* lives harmlessly, but sometimes this bacteria causes people to develop infections.
c. We want to understand why not all *Staphylococcus aureus* causes infections and why some *Staphylococcus aureus* is carried for longer than others.
d. To investigate this properly we need to look at the genetic material (DNA) in *Staphylococcus aureus*.

3. Why have I been chosen?

This General Practice is involved in the study with Oxford University Department of Primary Care and Oxford Radcliffe NHS Trust. We are inviting patients to join the study when booking

in for appointments. Your details are not known by anyone else other than the healthcare team at this practice. Only if you agree to take part in the study, would the research team be able to have your details. If you have any questions, you are welcome to speak to the study nurse or telephone a member of the study team whose number is at the top of this letter.

4. Do I have to take part?

It is up to you to decide whether or not to. This information sheet is to help you make this decision. If you do decide to take part you would be asked to sign a consent form. However, you will still be free to withdraw at any time without giving a reason.

5. What will happen to me if I take part?

If you are interested in participating, you could either arrange to meet the study nurse after your appointment or speak to a member of the study team whose number is at the top of this letter. If you agree, the study nurse will discuss the study with you and ask for your consent to take part. If you consent the nurse will take a sample from your nose using a swab, similar to a cotton bud, to test whether you are carrying *Staphylococcus aureus*. You will also be asked some questions about you and your family. We will also look at information that may be stored in your medical notes in the Oxford Radcliffe Hospital and at your GP practice.

We will send you a letter asking you to swab your nose yourself after one month and then again every two months for two years. Swabbing yourself is very simple and we will teach you the technique. Swabbing can be easily done at home with items we will send to you. Some people may find family members helpful in taking the swabs. The swab can be posted to us in stamped addressed envelopes that we will also give you. A study nurse will telephone you to remind you if you forget to send us the swab.

Participation in the study, or deciding not to participate, will not affect your medical treatment in any way.

6. What do I have to do?

Apart from considering whether to enter the study, you do not have to do anything else until you meet the study nurse. If you agree to take part, the nurse will ask you a few questions and explain how to take a nose swab yourself. Swabbing involves wiping the inside of your nose with a small cotton bud and sending it to the hospital. You will have to swab yourself, and provide information about antibiotics, after one month and then every two months for 2 years, 13 times in total. You will also need to provide us with consent to let us look at your GP records and hospital notes.

If you are carrying *Staphylococcus aureus* after two years, the nurse will contact you and ask you to carry on participating in the study (taking swabs from your nose every two months and posting them to us) for another two years.

7. What happens if I find I am carrying *Staphylococcus aureus?*

You may find out that you are carrying *Staphylococcus aureus* when you did not know this before. Many people carry *Staphylococcus aureus* and most never get infections. Part of this study is to better understand why many people carry these bacteria harmlessly in their nose and do not develop infections.

Methicillin resistant *Staphylococcus aureus* (MRSA)

MRSA is a strain of *Staphylococcus aureus* that has become resistant to some antibiotics. Finding out that you are carrying MRSA will not change your chances of developing infections. Neither will it affect your day to day activities. However, it may change the measures that nurses and doctors use when looking after you to try to prevent passing the bacteria onto other patients. For example, you may be looked after in a single room if you are admitted to hospital. If you need treatment with an antibiotic, it may affect the choice the doctor makes. If you are found to be carrying methicillin resistant *Staphylococcus aureus* (MRSA) your GP will be informed and a record will go onto your hospital records.

If you would like to know whether you are carrying *Staphylococcus aureus,* including methicillin resistant *Staphylococcus aureus,* from your samples please tick the box on your consent form and we will send you your previous result with each pack of swabs we send you.

8. What are the possible disadvantages and risks of taking part?

There is a very small risk of injury from self-swabbing if the cotton wool swab is pushed very far up the nose. The nurse will show you how to do this safely, and we will also send you a leaflet explaining how to swab each time we send you a swabbing kit.

Discovering that you are carrying methicillin resistant *Staphylococcus aureus* (MRSA) can cause people anxiety. Some people would prefer not to know this result and that is why you have the option of taking part in the study without being told your swab result.

9. What are the possible benefits of taking part?

There are no direct benefits to you from taking part in this study. We hope that by participating, you will help us find out more about how humans and *Staphylococcus aureus* affect each other which might benefit other patients in the future.

This completes Part 1 of the study information sheet if this has interested you and you are considering taking part in the study, please read the additional information in Part 2 before making a decision.

Part 2

10. What will happen if I don't want to carry on with the study?

If you wish to withdraw from the study we would like to use your data and samples up to your withdrawal but will not ask for any more. However, you can request that data or samples that have already been collected not be used. A decision to withdraw at any time, or a decision not to take part, will not affect your future medical care.

11. What if there is a problem or something goes wrong?

If you are harmed by taking part in this research project due to someone's negligence, then you may have grounds for a legal action but you may have to pay your legal costs. Regardless of this, if you wish to complain, or have any concerns about any aspect of the way you have been approached or treated during the course of this study, the normal National Health Service complaints mechanisms are available to you.

12. Will my taking part in this study be kept confidential?

Yes, all information collected about you during the course of the research will be kept strictly confidential. Your name will not be used on study forms, only a study code number, which will be kept separate from a secure list of participant names. Your name will not appear on any scientific reports or publications written as a result of this study. If you consent to take part nurses from the study will look at your hospital and GP records to get information about any visits in hospital and microbiology results during the course of the study. Authorised personnel from the Trust Research and Development may also look at your research records for purposes of audit.

13. What will happen to the samples and information I give?

The samples of *Staphylococcus aureus* and samples of its bacterial DNA will be stored for up to 10 years at the Nuffield Department of Medicine Laboratories in Oxford and will only be used for studies of *Staphylococcus aureus*. The answers to the questions the nurse asks you and details from your GP and medical notes are confidential, they will be kept securely and your name will not be in the same place as the answers you gave. After we have finished with the data the samples will be incinerated and your personal details securely disposed of.

14. What happens when the research study stops and what happens to the results?

At the end of the study the results will be made available to all doctors (through publication in medical journals). The references to these journals will be available on the study website http://www.medsci.ox.ac.uk/staph or alternatively you can email or write to us at the addresses at the top of this leaflet and we will send you a summary of the results of the study.

15. Who is organising and funding the research?

The study is a collaboration between the Nuffield Department of Medicine at the Oxford Radcliffe Hospital, Oxford and the Department of Primary Care at Oxford University. The

research is being funded by the National Institutes of Health Research Oxford Biomedical Research Centre. The National Institutes of Health Research are the research arm of the NHS.

16. Who has reviewed the study?

An independent group of people called a multi-centre Research Ethics Committee as well as external scientific experts have reviewed the study to protect your safety, rights, wellbeing and dignity.

17. Contact for further information

Your research nurse is the best person to ask for further information:

If the research nurses are not in the practice when you visit you can contact them by using the contact details at the top of this letter or phoning 01865 222884 and speak to them or leave a message. We will call you back as soon as possible to let you know when we will next be in the surgery.

If the research nurse is in the Surgery they will be very happy to see you.

You should keep a copy of this information sheet and you will be given a copy of the informed consent form.

Appendix 5: Carriage study consent form

Department of Primary Health Care

Oxford Radcliffe Hospitals NHS
NHS Trust

NDM Laboratory, Room 7724, Level 7, ORH Trust, Oxford, OX3 9DU
Tel: +44 (0)1865 222884, Fax: +44 (0)1865 222191
E-mail: staph@ndm.ox.ac.uk
Website: www.medsci.ox.ac.uk/staph

Study title: Population structure and carriage length of *Staphylococcus aureus*
Consent Form

Ethics committee number: 08/H0605/102
Chief Investigator: Dr Derrick Crook

Participant identification number:_____ *Please initial each box to show you agree:*

1.	I confirm that I have read and understood the information sheet dated 11/08/09 **(version 4.0)** for the above study and have had the opportunity to ask questions from ……and have had these answered satisfactorily.	
2.	I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected.	
3.	I agree that the nurse can swab my nose now, to be used as part of the research study	
4.	I agree to swab my nose and send off the swab in stamped addressed envelopes provided after one month and again every two months for two years. During this time I also agree to record any antibiotics I take not prescribed to me by my GP on the forms provided.	
5.	I agree if *Staphylococcus aureus* is isolated from a clinical sample from me, whilst I am in hospital, or during the next two years it can be used for research.	
6.	I understand that if my samples grow *Staphylococcus aureus* bacteria it will be stored as part of this study and used in future research.	
7.	I understand that my GP will be contacted members of the research team and my GP records will be looked at, at the beginning of the study and again after one and two years. I give permission for these individuals to contact my GP and have access to my GP records.	
8.	I understand that sections of my medical notes may be looked at by members of the research team. I give permission for these individuals to have access to my medical records.	
9.	I understand that I may be contacted again after two years and asked to continue participation in the study.	
10.	I would like to be informed of the results of my nasal swab during the study	

I agree to take part in the above study

_____	_____	_____
Name of patient	Date	Signature
_____	_____	_____
Name of researcher	Date	Signature

When completed, 1 for patient, 1 for research staff; 1 (original) to be kept in medical notes

Appendix 6: Carriage study participant and GP record questionnaires

ID | | | | |

S. aureus	

Population structure and carriage length of *Staphylococcus aureus* in Oxfordshire

Questionnaire

ID				

Form 2: Personal details *Confidential*

1. Name

Surname:_____
Forename(s):_____

2. Date of birth

Day	*Month*	*Year*

3. Sex

Male ☐ Female ☐

4. Hospital Number

5. NHS Number

6. **Recruitment Group**

7. Address

Postcode:

8. **Contact numbers**

	Preferred number
Home:	☐
Work:	☐
Mobile:	☐

9. Usual GP name

Surname:_____
Initials:_____

10. GP practice address

Postcode:

11. GP Contact number

12. Person undertaking interview

Interviewer number

ID | | | | |

Form 3: Participant interview

13. Date of interview and baseline swab

Day	*Month*	*Year*

14. Date of Birth / /

15. Sex Male ☐ Female ☐

16. Ethnic background
Please tick one box

White	*British*	
	Irish	
	Any other white background	
Mixed	*White and Black Caribbean*	
	White and Black African	
	White and Asian	
	Any other Mixed background	
Asian or Asian British	*Indian*	
	Pakistani	
	Bangladeshi	
	Any other Asian Background	
Black or Black British	*Caribbean*	
	African	
	Any other Black background	
Chinese or other ethnic group	*Chinese*	
	Any other	
Does not wish to give ethnic group		
Not stated		

17. Are you currently in employment?

Yes		***Please go to question 17***
No		***Please go to question 18***

18. What is your current employment?

19. Have you ever had any healthcare related employment?

Yes		*Please describe below*
No		*Please go to question 19*

Date of most recent healthcare employment	Year:	Month:
Nature of last healthcare related employment		

ID				

Household details

20. Do you live alone? Refer to usual household

Yes		*Go to question 21*
No		*Please complete details below*

21. Do you live in any of the following? Refer to usual household.

Residential home		
Nursing home		
Other medical or care home		
Defence establishment		
Prison Service Establishment		
Educational establishment (including halls of residence)		
Hotel, boarding house or guest house		
Hostel		
None of above / Own home		If own home / none of the above please give details of your household in the table below, including children/teenagers

Sex	*Age*	*Occupation*	*Does the household member have any frequent healthcare contact e.g. visits to hospital or GP or care from nurses in the community.* Yes	No	Details

22. Do you currently take part in any gym / sports activities (involving contact sport or shared changing facilities)?
Yes: ☐ No ☐
If yes – times per month: _____

ID | | | | |

23. Have you ever travelled outside of the UK? No:☐ Yes☐ If yes:
When did you last travel abroad?

Year:	Month:

Which country did you last travel abroad to:_____
Have you been abroad to any of the following continents in the last year?

Continent	***Yes/No***
Europe	*Yes* ☐ *No* ☐
North America	*Yes* ☐ *No* ☐
Central and South America	*Yes* ☐ *No* ☐
Africa	*Yes* ☐ *No* ☐
Asia	*Yes* ☐ *No* ☐
Australia	*Yes* ☐ *No* ☐

24. Do you look after, or give any help or support to family members, friends, neighbours or others because of:
Do not include anything you do as part of your paid employment.

Long term physical or mental ill health or disability	*Yes* ☐ *No* ☐
Problems related to old age	*Yes* ☐ *No* ☐

25. Do you receive care, other than nursing care, at home?

Type of care	*Yes/No*	Number of visits per day	Number of days per week
Personal care e.g. help with dressing, taking medicines, washing or bathing	*Yes* ☐ *No* ☐		
Domestic help e.g. help with housework or cooking	*Yes* ☐ *No* ☐		
Other social service support e.g. meals on wheels, help with shopping	*Yes* ☐ *No* ☐		

26. Have you ever undergone any of the following types of care of procedures?
If yes please give date of most recent procedure

Procedure:	Yes/No	Year	Month
Vascular access	*Yes* ☐ *No* ☐		
Insertion of urinary or suprapubic catheter	*Yes* ☐ *No* ☐		

Interviewer's signature	Print name	Date

ID				

Form 4 Information from GP records

27. Date of interview and baseline swab *Day* *Month* *Year*

28. Date of Birth ____/____/____

29. Sex Male ☐ Female ☐

30. Has the participant ever had any hospital in-patient episodes outside of ORH hospitals (JR, Churchill, Radcliffe Infirmary, Horton)?
No☐ Yes ☐ If yes:
Please give the date of last in-patient episode outside of ORH hospitals:
____/____/____
Complete table for admissions outside ORH in year prior to baseline swab and print out first page of discharge letter for each admission.

Date of admission	*Date of discharge*	*Hospital*	*Discharge Diagnosis*

31. Has the participant ever had any hospital out-patient episodes outside of ORH hospitals (JR, Churchill, Radcliffe Infirmary, Horton)?
No☐ Yes ☐ If yes:
Please give date of last out-patient episode outside ORH hospitals: ____/____/____
Complete table for out-patient episodes outside ORH in year prior to baseline swab.

Date of outpatient attendance	*Hospital*

32. Has the participant ever had an appointment with their GP?
No☐ Yes ☐ If yes:
Please give date of last GP appointment: _____/_____/_____
Complete table for GP appointments in year prior to baseline swab.

Date of GP appointment

ID				

33. Has the participant ever had any appointments with the practice nurse?
No☐ Yes ☐ If yes:
Please give date of last practice nurse appointment: _____/_____/_____
Complete table for practice nurse appointments in year prior to baseline swab.

Date of appointment	*Reason*	*Did this involve dressings, ulcer or wound care?*
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐

34. Has the participant ever received community / district nurse or visiting care?
No☐ Yes ☐ If yes:
Please give date of last district nurse appointment: _____/_____/_____
Complete table for community / district nurse or visiting care appointments in year prior to baseline swab.

Date of appointment	*Reason*	*Did this involve dressings, ulcer or wound care?*
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐
		Yes☐ No☐

35. According to the general practice problem list, does the participant have any long term illness?

Yes		*Please print problem list*
No		*Go to Question 34*

36. Has the participant ever undergone any of the following types of care of procedures?

Procedure	Yes/No	Date of most recent procedure	Approx.
Chemotherapy	*Yes* ☐ *No* ☐		☐
Renal dialysis	*Yes* ☐ *No* ☐		☐
Surgery	*Yes* ☐ *No* ☐		☐
Prescription of oral steroids	*Yes* ☐ *No* ☐		☐

ID				

37. In the past month, or currently, has the participant received treatment for:
    - Trauma/Cut/Splinter
    - Cannula
    - Arterial or venous catheter
    - Surgery
    - Eczema/Psoriasis
    - Skin break down - Ulcer/Rash (e.g. chicken pox, herpes, zoster)
    - Insect bite/Bite
    - Tattoo/Piercing

    Yes: ☐ No ☐

    If yes – details: _____

38. In the past month, or currently, has the participant had percutaneous vascular access?
    Yes: ☐ No ☐

39. Has the participant ever received antimicrobials?
    No☐ Yes ☐ If yes:
    Please give date of last antimicrobial prescription: / /
    Complete table for antimicrobials prescribed in year prior to baseline swab.

Antimicrobial	*Date*	*Dose*	*Number of Tablets*

40. For patients recruited in hospital, has the participant received antimicrobials while an in-patient?
    Yes ☐ No☐ Not applicable ☐
    If yes, complete table

Antimicrobial name	*Date course started*	*Date course ended*

41. Has MRSA been isolated previously?

Yes		*Enter details below*
No		

Date of first isolation	
Type of sample	

42. Has MSSA been isolated previously?

Yes		*Enter details below*
No		

Date of first isolation	
Type of sample	

Comments:

_____
_____
_____

Interviewer's signature	Print name	Date

End of questionnaire

Appendix 7: Carriage study swab form

Department of Primary Health Care

Oxford Radcliffe Hospitals NHS
NHS Trust

NDM Laboratory, Room 7724, Level 7, ORH Trust, Oxford, OX3 9DU
Tel: +44 (0)1865 222884, Fax: +44 (0)1865 222191
E-mail: staph@ndm.ox.ac.uk
Website: www.medsci.ox.ac.uk/staph

Study title: Population structure and carriage length of *Staphylococcus aureus*

Ethics committee number: 08/H0605/102

Chief Investigator: Dr Derrick Crook

Form F1 **Participant ID_____**

Please complete the details in the form below about the date you took your nasal swab and antibiotics you have been taking and send the form back to us with your swab packs.

Date nasal swab taken _____

Name of antibiotic	**Date you started the antibiotic**	**Date you finished the antibiotic / whether you are still taking it**	**Amount you take per day**
Example	08.09.2008	Still taking	50 mg twice daily

Appendix 8: Carriage study instructions for taking a nasal swab

UNIVERSITY OF OXFORD

Department of Primary Health Care
Oxford Radcliffe Hospitals NHS
NHS Trust

NDM Laboratory, Room 7724, Level 7, ORH Trust, Oxford, OX3 9DU
Tel: +44 (0)1865 222884, Fax: +44 (0)1865 222191
E-mail: staph@ndm.ox.ac.uk
Website: www.medsci.ox.ac.uk/staph

Population structure and carriage length of *Staphylococcus aureus*
Ethics committee number: 08/H0605/102
Chief Investigator: Dr Derrick Crook

How to take a nasal swab

Please find enclosed the next swab pack for the study "Population structure and carriage length of *Staphylococcus aureus*" in which you are participating.
The swab pack should contain:

- The result from your previous swab, if you requested it
- Antibiotic information form F1
- One nose swab and tube
- Protective packaging for swab and container (stiff plastic case and clear plastic bag)
- Postage paid envelope for returning the swab and form

Please use the swab in the pack provided to take a swab from your nose. Please also fill in the attached form asking for details about antibiotics you may be taking. A reminder on how to take a swab is found overleaf.
Once you have taken the swab and completed the information form please send them back to us in the postage paid envelope provided.

If you:

- have any questions or concerns
- would like a further explanation about how to take a nasal swab or complete the antibiotics form
- have anything missing from your pack
- are worried about your results, which may have been provided with this pack if you requested them

Either see our website: www.medsci.ox.ac.uk/staph OR contact us by:
Phone: 01865 222884
Email: staph@ndm.ox.ac.uk

UNIVERSITY OF OXFORD

Department of Primary Health Care
Oxford Radcliffe Hospitals NHS
NHS Trust

NDM Laboratory, Room 7724, Level 7, ORH Trust, Oxford, OX3 9DU
Tel: +44 (0)1865 222884, Fax: +44 (0)1865 222191
E-mail: staph@ndm.ox.ac.uk
Website: www.medsci.ox.ac.uk/staph

Population structure and carriage length of *Staphylococcus aureus*
Ethics committee number: 08/H0605/102
Chief Investigator: Dr Derrick Crook

How to take a nasal swab

Please find enclosed the next swab pack for the study "Population structure and carriage length of *Staphylococcus aureus*" in which you are participating.
The swab pack should contain:

- The result from your previous swab, if you requested it
- Antibiotic information form F1
- One nose swab and tube
- Protective packaging for swab and container (stiff plastic case and clear plastic bag)
- Postage paid envelope for returning the swab and form

Please use the swab in the pack provided to take a swab from your nose. Please also fill in the attached form asking for details about antibiotics you may be taking. A reminder on how to take a swab is found overleaf.
Once you have taken the swab and completed the information form please send them back to us in the postage paid envelope provided.

If you:

- have any questions or concerns
- would like a further explanation about how to take a nasal swab or complete the antibiotics form
- have anything missing from your pack
- are worried about your results, which may have been provided with this pack if you requested them

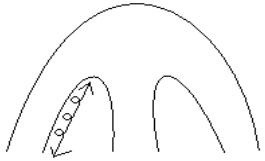
Either see our website: www.medsci.ox.ac.uk/staph OR contact us by:
Phone: 01865 222884
Email: staph@ndm.ox.ac.uk

Taking the swab

Taking a nasal swab is very simple. You will have been shown the technique by the study nurse when at your first meeting.
Some people may find it helpful to ask a friend or family member to help take the swab, but this is not always necessary.

The steps to take a swab are shown below:

1. Open the packet and remove the swab—take care not to touch the end of the swab with your fingers.
2. Place the swab inside the tip of your nose—you do no have to put it far up the nose, just enough to cover the cotton tip of the swab
3. Pressing lightly, move the swab up and down the front of the inside of your nose three times. Whilst you are doing this turn the swab completely round (360º) in your nose three times clockwise and again three times anti-clockwise



Nose from beneath

WARNING !!!!
Placing the swab too far up the nose may cause injury. Please only place the cotton tip of the swab in the nostril

4. Repeat this for both nostrils using the same swab
5. Place the swab in its tube
6. Complete form F1 (see overleaf for instructions)
7. Place the swab and tube in the stiff plastic protective packaging and close securely.
8. Place the protective packaging in the clear plastic bag
9. Place the above, along with form F1, in the postage paid envelope and seal
10. Send off in the postage paid envelope. Please post the swab as soon as possible after taking it.

Completing the antibiotics form

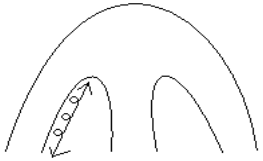
- We are also asking you to fill in the enclosed form with details about antibiotics you may be taking at the moment, or have taken in the last 2 months. We are interested in antibiotics that were **not** prescribed by your GP, for example antibiotics you may have been prescribed by your dentist, by another clinic you attend or when you were abroad.
- The form (F1) is provided in this pack with the swabbing kit.
- On the form please record the name of any antibiotics you are taking that were **not** prescribed to you by your doctor. The form asks for details about how often you take it, in what dose (amount) and the dates you were taking it from and to (or that you are still taking it).
- If you are not sure whether something you are taking is an antibiotic, please write down the name of the medication and fill in the details anyway.
- Please also record the date you took your nasal swab on the form.

Taking the swab

Taking a nasal swab is very simple. You will have been shown the technique by the study nurse when at your first meeting.
Some people may find it helpful to ask a friend or family member to help take the swab, but this is not always necessary.

The steps to take a swab are shown below:

1. Open the packet and remove the swab—take care not to touch the end of the swab with your fingers.
2. Place the swab inside the tip of your nose—you do no have to put it far up the nose, just enough to cover the cotton tip of the swab
3. Pressing lightly, move the swab up and down the front of the inside of your nose three times. Whilst you are doing this turn the swab completely round (360º) in your nose three times clockwise and again three times anti-clockwise



Nose from beneath

WARNING !!!!
Placing the swab too far up the nose may cause injury. Please only place the cotton tip of the swab in the nostril

4. Repeat this for both nostrils using the same swab
5. Place the swab in its tube
6. Complete form F1 (see overleaf for instructions)
7. Place the swab and tube in the stiff plastic protective packaging and close securely.
8. Place the protective packaging in the clear plastic bag
9. Place the above, along with form F1, in the postage paid envelope and seal
10. Send off in the postage paid envelope. Please post the swab as soon as possible after taking it.

Completing the antibiotics form

- We are also asking you to fill in the enclosed form with details about antibiotics you may be taking at the moment, or have taken in the last 2 months. We are interested in antibiotics that were **not** prescribed by your GP, for example antibiotics you may have been prescribed by your dentist, by another clinic you attend or when you were abroad.
- The form (F1) is provided in this pack with the swabbing kit.
- On the form please record the name of any antibiotics you are taking that were **not** prescribed to you by your doctor. The form asks for details about how often you take it, in what dose (amount) and the dates you were taking it from and to (or that you are still taking it).
- If you are not sure whether something you are taking is an antibiotic, please write down the name of the medication and fill in the details anyway.
- Please also record the date you took your nasal swab on the form.