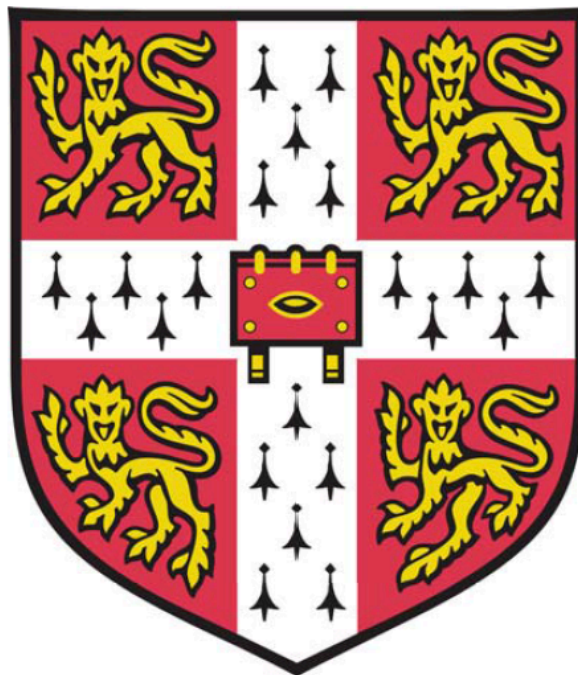**Genomic surveillance of methicillin-resistant *Staphylococcus aureus* at local, regional and national levels**

Michelle Suzanne Toleman

Darwin College
Department of Medicine
Wellcome Sanger Institute
University of Cambridge

June 2019

This dissertation is submitted for the degree of
Doctor of Philosophy

Declaration

This thesis describes work carried out between January 2015 and June 2019 under the supervision of Professor Sharon Peacock at the Department of Medicine, University of Cambridge and Professor Julian Parkhill at the Wellcome Sanger Institute.

This dissertation is the result of my own work. Any work that is the outcome of collaboration is indicated in the collaborative contributions section of the text.

This work is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as specified in the text. No substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as specified in the text.

This work contains less than 60,000 words.

Michelle Suzanne Toleman, June 2019

Genomic surveillance of methicillin-resistant *Staphylococcus aureus* at local, regional and national levels

Michelle S. Toleman

Abstract

Methicillin-resistant *Staphylococcus aureus* (MRSA) is a major cause of hospital-acquired infection, which is associated with increased cost and length of hospital stays and substantial morbidity and mortality. The rapid progression of whole-genome sequencing (WGS) technologies over the past decade has provided highly detailed and discriminatory insights into the epidemiology of MRSA. However, WGS is still not implemented in the routine surveillance of MRSA at a local, regional or national level. This thesis applies WGS to a number of questions and populations to describe its benefits.

Standard infection control investigation commonly uses the antimicrobial susceptibility patterns (ASPs, patterns of susceptibility and resistance to commonly used antibiotics) of MRSA as a surrogate for bacterial relatedness. As they are readily available, ASPs are often combined with patient movement to evaluate putative MRSA outbreaks in hospitals. The accuracy of this method was evaluated by comparing linked cases based on MRSA ASPs versus linked cases based on whole-genome relatedness, using data from a year-long prospective observational cohort study of 1,465 MRSA-positive individuals. The sensitivity and specificity of ASP in the presence of a direct ward contact was 44% and 85%, respectively; in the presence of a shared residential post code, the sensitivity and specificity of ASP is 59% and 76%, respectively. This demonstrates that compared to WGS plus epidemiology, ASP and epidemiology does not reliably identify or refute transmission events.

A lineage referred to as epidemic (E)-MRSA15 is largely considered to be associated with hospital settings, but an epidemiological and genomic investigation of a MRSA outbreak in a General Practice (GP) identified 15 people who were E-MRSA15 positive, the majority of which shared a link to a leg ulcer/podiatry clinic in the GP surgery. The outbreak had not been detected previously, and was only identified post-hoc from MRSA sequence data, highlighting the importance of MRSA sequencing to detect otherwise cryptic community outbreaks.

The utility of WGS for the investigation of regional MRSA epidemiology was explored for two potentially high-risk MRSA lineages (USA300 and ST2371) that are otherwise unmonitored by current surveillance. Screening the genomes of MRSA isolated from 1,465 people identified over a 12-month period demonstrated that 4.2% of cases were positive for MRSA USA300, with multiple introductions and household transmissions identified. Five people were positive for ST2371, all of whom had a direct or indirect link to a substantial outbreak in an intensive care unit at Addenbrooke's Hospital in 2011, thus confirming the value of WGS in regional epidemiological investigations.

The feasibility and utility of incorporating WGS into routine national MRSA surveillance was evaluated through a combined epidemiological and genomic survey of MRSA bacteraemia undertaken in England over a one-year period. This captured 903 reported cases of MRSA bacteraemia, with 425 isolates available for sequencing. Almost two thirds of isolates were assigned to multi-locus sequence type clonal complex (CC) 22. The addition of MRSA genomes from published outbreak investigations showed that the study genomes could provide context for outbreak isolates and supported cluster identification. Potentially high-risk lineages were also detected. These findings support the integration of epidemiological and genomic surveillance for MRSA bacteraemia as a first step towards a comprehensive national surveillance programme.

Acknowledgements

Collaborative contributions

The collaborative contributions to this work are listed by chapter.

Chapter 3.

Francesc Coll provided data from the Cambridgeshire prospective MRSA study. I undertook the analysis and interpretation of data. Sandra Reuter, Julian Parkhill and Emmeline Watkins provided valuable guidance. Sharon Peacock and Ewan Harrison supervised the study.

Chapter 4.

USA300 study: Beth Blane conducted the laboratory work. Nicholas Brown provided clinical laboratory support. I undertook the phylogenetic and epidemiological analysis of data. Ewan Harrison and Sandra Reuter provided valuable guidance. Francesc Coll generated Figure 4.1. Sharon Peacock, Estée Török and Julian Parkhill supervised the study.

ST2371 study: Beth Blane conducted the laboratory work. Francesc Coll undertook the phylogenetic analysis as a training aid, and generated Figure 4.5 and 4.6. I collected and analysed patient level epidemiological and clinical data. Ewan Harrison and Sandra Reuter provided valuable advice. Sharon Peacock supervised the study.

Chapter 5.

Outbreak meetings were chaired by Nicholas Brown and were regularly attended by myself alongside Emmeline Watkins, Belinda Sadler, Bernadette Nazareth, Rachel Thaxter and the practice manager. Beth Blane cultured isolates and extracted DNA. I carried out sampling of staff and the environment on-site with Belinda Sadler. Environmental samples were processed in the clinical microbiology laboratory at CUH. Bernadette Nazareth collated the patient information sheet and consent form. Patient data was extracted by the practice manager. Epidemiological analysis was undertaken by myself and Emmeline Watkins. I conducted the bioinformatic analysis, interpreted the data and generated the figures. Francesc Coll, Ewan Harrison, Jill Eastment, Lynn Rodrigues, Estée Török, Sandra

Reuter and Dorota Jamrozy provided valuable advice. Sharon Peacock and Julian Parkhill supervised this work.

Chapter 6.

Beth Blane and Hayley Wilson performed the laboratory work in Cambridge. Russell Hope and Angela Kearns co-ordinated the collection and characterisation of isolates and provided the epidemiological and laboratory data from PHE Colindale. Francesc Coll provided data from the Cambridgeshire MRSA prospective study. I performed epidemiological and bioinformatic analyses and produced the figures. Assistance was provided by Sandra Reuter, and from Dorota Jamrozy regarding *in silico spa*-typing. Ewan Harrison, Sharon Peacock and Julian Parkhill provided valuable advice. Estée Török supervised this work.

Table of Contents

Tables

Figures

List of Abbreviations

ACME	Arginine catabolic mobile element
ASP	Antibiotic susceptibility profile
BSAC	British Society of Antimicrobial Chemotherapy
CA-MRSA	Community-associated MRSA
CC	Clonal complex
CCG	Clinical commissioning group
ccr	Cassette chromosome recombinase
COMER	Copper and mercury resistance element
CUH	Cambridge University Hospitals Foundation NHS Trust
DNA	Deoxyribonucleic acid
DST	Direct susceptibility testing
E	Epidemic
EARs-NET	European Antimicrobial Resistance Surveillance Network
ECDC	European Centre for Disease Control
ED	Emergency Department
ENA	European Nucleotide Archive
EUCAST	European Committee on Antimicrobial Susceptibility Testing
GP	General Practice
ICT	Infection control team
IT	Information technology
IQR	Interquartile range
LIMS	Laboratory information management system
LTCF	Long-term care facility
Maldi-TOF	Matrix-assisted laser desorption ionisation time of flight mass spectroscopy
MESS	Mandatory enhanced surveillance scheme
MGE	Mobile genetic element
MLST	Multi-locus sequence typing
MRSA	Methicillin-resistant *Staphylococcus aureus*
MSSA	Methicillin susceptible *Staphylococcus aureus*
NHS	National Health Service
PBP	Penicillin-binding protein

PFGE	Pulsed-field gel electrophoresis
PCR	Polymerase chain reaction
PHE	Public Health England
PIR	Post infection review
PVL	Panton-Valentin Leukocidin
PPV	Positive predictive value
SAPI5	*S. aureus* pathogenicity island 5
SARs-CoV	Severe Acute Respiratory Syndrome Coronavirus
SCC*mec*	Staphylococcal cassette chromosome *mec*
SCBU	Special care baby unit
SGSS	Second Generation Surveillance System
spa	*Staphylococcus aureus* Protein A
SNP	Single nucleotide polymorphism
SSTI	Skin and soft tissue infection
ST	Sequence type
US	United States of America
UK	United Kingdom
WGS	Whole-genome sequencing
WSI	Wellcome Sanger Institute

Publications from this Thesis

Local persistence of novel MRSA lineage after hospital ward outbreak, Cambridge, UK, 2011–2013 [letter]. **Toleman MS**, Reuter S, Coll F, Harrison EM, Peacock SJ. *Emerging Infectious Diseases.* 2016 Aug. doi: 10.3201/eid2209.151100.

Systematic surveillance detects multiple silent introductions and household transmission of methicillin-resistant *Staphylococcus aureus* USA300 in the east of England. **Toleman MS**, Reuter S, Coll F, Harrison EM, Blane B, Brown NM, Török ME, Parkhill J, Peacock SJ. *Journal of Infectious Diseases.* 2016 Apr. doi: 10.1093/infdis/jiw166.

Investigation of a Cluster of Sequence Type 22 Methicillin-resistant *Staphylococcus aureus* Transmission in a Community Setting. **Toleman MS**, Watkins ER, Williams T, Blane B, Sadler B, Harrison EM, Coll F, Parkhill J, Nazareth B, Brown NM, Peacock SJ. *Clinical Infectious Diseases.* 2017 Nov 29;65(12):2069-2077. doi: 10.1093/cid/cix539.

Genomic Surveillance of MRSA associated with bloodstream infection in England. **Toleman MS**, Reuter S, Jamrozy D, Wilson HJ, Blane B, Harrison EM, Coll F, Hope RJ, Kearns A, Parkhill J, Peacock SJ, Török ME. *Eurosurveillance.* 2019; 24(4):pii=1800215. https://doi.org/10.2807/1560-7917.ES.2019.24.4.1800215

Other relevant publications

Genomic surveillance reveals low prevalence of livestock-associated methicillin-resistant *Staphylococcus aureus* in the East of England. Harrison EM, Coll F, **Toleman MS**, Blane B, Brown NM, Török ME, Parkhill J, Peacock SJ. *Scientific Reports*. 2017 Aug 7;7(1):7406. doi: 10.1038/s41598-017-07662-2.

Longitudinal genomic surveillance of MRSA in the UK reveals transmission patterns in hospitals and the community. Coll F, Harrison EM, **Toleman MS**, Reuter S, Raven KE, Blane B, Palmer B, Kappeler ARM, Brown NM, Török ME, Parkhill J, Peacock SJ. *Science Translational Medicine*. 2017 Oct 25;9(413). pii: eaak9745. doi: 10.1126/scitranslmed.aak9745.

Chapter 1. Introduction

1.1. *Staphylococcus aureus*: Microbiology

1.1.1. Microbiological Characteristics and Identification

Staphylococcus aureus is a facultative anaerobic Gram-positive coccus, which is typically distinguished microbiologically from other staphylococci by its gold coloured colonies on blood agar and production of the enzyme, coagulase. It ferments mannitol, produces DNAse and catalase, and tends to appear beta-haemolytic on blood agar. It is a hardy non-sporulating bacterium which tolerates high salinity, temperatures up to approximately 60ºC, and resists drying.[1]

1.1.2. Colonisation and Disease

A human commensal, the most common site of colonisation of *S. aureus* in adults is the anterior nares.[2,3] Approximately 20% of individuals are persistent nasal carriers, 30% intermittent carriers and 50% non-carriers.[3] Environmental, host and bacterial factors have been shown to impact the colonisation status of an individual. Rates of colonisation are high in those undertaking dialysis, those with diabetes, those undergoing surgery and those with immunodeficiencies.[4]

S. aureus is one of the most common causes of bacterial infection and can cause disease both locally through infection of the a body site, or at a distance through toxin production.[5] If *S. aureus* is isolated from a site manifesting disease (showing clinical signs of infection) or a sterile site of the body, such as blood or joint fluid, then the microbiological diagnosis

of infection is made. Asymptomatic carriage is a strong contributor to the risk of disease because infection with *S. aureus* tends to be endogenous in nature.[2,6] Persistent nasal carriage is also associated with a higher risk of infection.[7] It is a common cause of skin and soft tissue infection (SSTI) but can cause a wide range of infection from uncomplicated SSTI such as furuncles and carbuncles to serious invasive diseases such as infective endocarditis and osteomyelitis. An increased risk of infection is seen in those who are male, young or elderly.[5]

1.1.3. Methicillin Resistance

In day-to-day clinical practice, *S. aureus* is commonly referred to as methicillin-susceptible (MSSA), or methicillin-resistant (MRSA), as this impacts upon treatment regimens. Methicillin is no longer used clinically, and in the United Kingdom (UK), the β-Lactam antibiotic flucloxacillin is a typical first-line treatment for *S. aureus* infections. MSSA is susceptible to the majority of β-Lactam antibiotics. MRSA, however, is resistant to most β-Lactam antibiotics, including penicillins, cephalosporins and carbapenems. This includes the commonly used semi-synthetic penicillin, flucloxacillin. Resistance is due to the presence of a mobile genetic element (MGE) known as the staphylococcal cassette chromosome *mec* (SCC*mec*) within the staphylococcal chromosome.[8,9] The presence of SCC*mec* is therefore sometimes considered the defining feature of MRSA.[10]

MRSA was first isolated in the UK in 1961.[11] Based on epidemiological evidence, it had long been thought that the introduction of methicillin into widespread clinical practice in 1960 was the main evolutionary driver for MRSA emergence. However, a study by Harkins and colleagues in 2017 showed using Bayesian phylogenetic reconstruction that an

ancestral type I SCC*mec* element was acquired by a methicillin-susceptible strain of *S. aureus* in the mid-1940s.[12] They hypothesised, therefore, that the widespread use of first generation β-Lactam antibiotics such as penicillin was the main driver for widespread dissemination of a pre-existing lineage of MRSA.

Currently, there are 11 types of SCC*mec* element recognised in *S. aureus*, each consisting of a unique combination of *mec* gene complex and the cassette chromosome recombinase (*ccr*) gene complex.[10,13] The *mec* gene complex carries one of two genes which, to date, have been recognised to encode methicillin resistance. In the vast majority of MRSA lineages, the *mec* gene complex contains *mecA*, a gene which encodes an altered penicillin binding protein (PBP) 2a.[14,15] The bacteriostatic action of β-Lactam antibiotics is mediated by competition for the active site of the transpeptidase domain of PBPs. This blocks access to the substrate and prevents the cross-linking of the peptidoglycan strands that is required for efficient bacterial cell wall synthesis.[16] Therefore, because MRSA has an altered PBP, it has a lowered affinity to β-Lactam antibiotics and cell wall synthesis can continue despite the antibiotic presence.[15]

There are, however, some exceptions to the descriptions above. A minority of MRSA isolates do not carry the *mecA* gene but instead the *mec* complex contains a *mecC* gene. Reported in 2011, this *mecA* homologue encodes a modified PBP2a to mediate methicillin resistance.[17] A recent report by Becker *et al.*, 2018, describes the first case of a *mecB* gene being identified in an MRSA from a nasal screening sample in Muenster, Germany. Prior to this, the *mecB* gene had only been identified in *Macrococcus caseolyticus,* not within staphylococcal species.[18] In addition, whilst it is still terminology in common use, MRSA is no longer considered resistant to all β-Lactam antibiotics. In the last decade a number of

altered β-Lactam antibiotic agents that maintain some activity against MRSA have been developed, such as Ceftaroline and Ceftobiprole.[19,20]

1.2. Methicillin-resistant *Staphylococcus aureus*

1.2.1. Clinical Features of MRSA Infection

MRSA causes the same broad spectrum of disease as that seen in MSSA infections (Section 1.1.2). These broad clinical manifestations mean that clinical suspicion for MRSA infection is often based on the identification of risk factors. The people at greatest risk for MRSA infection are those with a history of prolonged hospital admission, surgery (within a year), renal dialysis, indwelling device/catheter or residence within a long-term care facility (LTCF).[21] Whilst MRSA is not necessarily more virulent then susceptible strains, MRSA infections have been shown to be associated with an increased length of stay in hospital, increased mortality and higher costs.[22] There are a number of factors hypothesised to contribute to this, such as increased time to optimal antibiotic therapy being commenced and presence of co-morbidities and complications in the hospital populations where infection rates are highest.[23,24]

1.2.3. History of MRSA in England

Currently, successful lineages of MRSA are widely disseminated globally, especially within hospitals. Stefani *et al.*, 2012, reviewed a number of global studies and describe how MRSA rates greater than 50% were seen in North and South America, Asia and Malta. The lowest prevalence rates are seen in Scandinavia and the Netherlands.[25] Throughout history, *S. aureus* lineages have been observed to emerge, expand, and then decline,[26] and MRSA

appears to behave no differently.[27] Around the time of its discovery in the early 1960s, MRSA was uncommon in England.[28] There was a marked increase in MRSA infections in the 1980s with the spread of an MRSA lineage known as epidemic (E) EMRSA-1.[29] Reacher *et al.*, 2000, described how methicillin-resistance increased from 1.7% to 3.8% of all *S. aureus* from 1990 to 1993, to 32% in 1997 and reached 34% in 1998.[30] The increasing infection rates generated a high level of public concern and media attention. Rates of MRSA bacteraemia reached an all-time high by 2003 with around 7,700 cases per year,[31] followed by a gradual decrease to current levels. Since 2014, levels have remained relatively constant at around ~800 cases of MRSA bacteraemia per year.[32]

Whilst the rise in MRSA infection in the 1990s – 2000s has been attributed to the expansion of EMRSA-16 and EMRSA-15 lineages,[33,34] it is more difficult to disentangle the cause of the decline of MRSA in England. Historic decreases, such as those seen after early outbreaks in Switzerland and Denmark in the 1970s, were attributed to improved infection control and antibiotic stewardship.[35,36] MRSA bacteraemia in England declined at a substantial rate alongside control initiatives such as active surveillance, isolation, decolonisation regimens, improved care of intravenous devices and environmental decontamination being introduced. Due to such a rapid introduction of these initiatives, it has not been possible to robustly determine the contribution of each intervention to the decline in rates. Indeed, it has been proposed that one of the endemic lineages, EMRSA-16, was in fact already in decline prior to these initiatives being introduced.[37] Conflicting with the hypothesis that the decline of MRSA bacteraemia is attributed to strengthened infection control initiatives is that the rate of MSSA bacteraemia, which should respond in a similar way to such interventions, has continued to increase, rather than decrease, since rates began to be monitored in 2011.[32] Because of this uncertainty, it is commonly regarded

that it is most likely that a multifaceted strengthening of infection control measures alongside potential changes in the population structure of MRSA have together caused the observed reduction in MRSA bacteraemia.[37-39]

1.2.4. Current Epidemiology of MRSA Bacteraemia in England

Figure 1.1 demonstrates the substantial variation in the distribution of MRSA across Europe, with a North-South gradient visible. Comparison between the two maps shows the substantial decrease in MRSA in England in the last decade. The most recent report shows that England reported about 5-10% of invasive *S. aureus* isolates as methicillin-resistant in 2016.
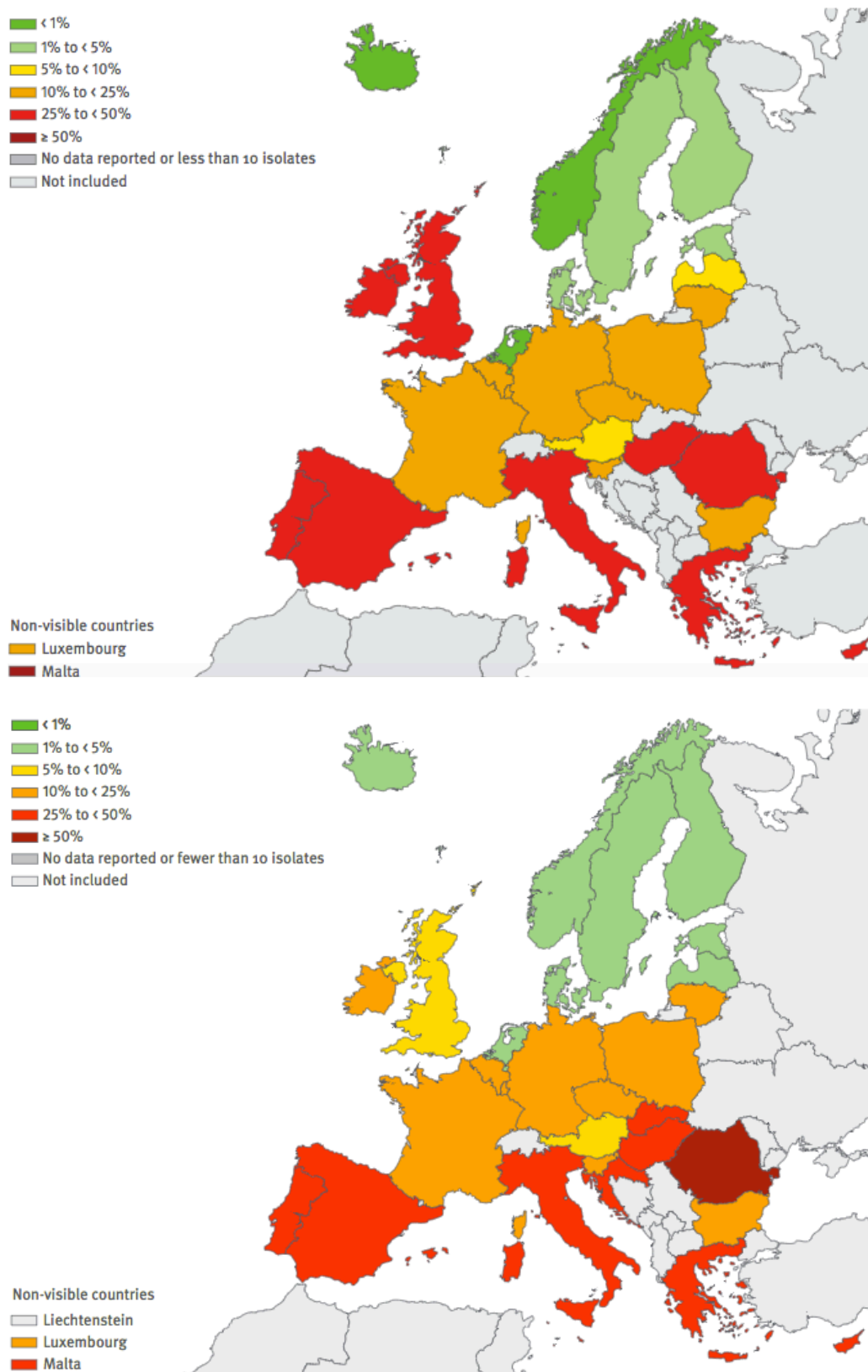
Figure 1.1. Proportion of invasive *Staphylococcus aureus* resistant to methicillin, MRSA, by EU/EEA countries, 2009 and 2016. Reproduced from: Surveillance of antimicrobial resistance in Europe, ECDC, 2009 and 2016.

Within England, absolute numbers of MRSA bacteraemia cases collected through mandatory reporting form the primary method of surveillance (Section 1.5.3). Over the last five years, rates of MRSA bacteraemia have remained at around 1.5 per 100,000 population per year. Surveillance indicates that there is geographical variation in the rates of MRSA bacteraemia across England with a North-South trend (Figure 1.2). NHS England North (Greater Manchester) report the highest rates at 2.1 cases per 100,000 in 2016/2017, with NHS England South (Wessex) reporting the fewest cases at 0.7 cases per 100,000 in the same year.[32]
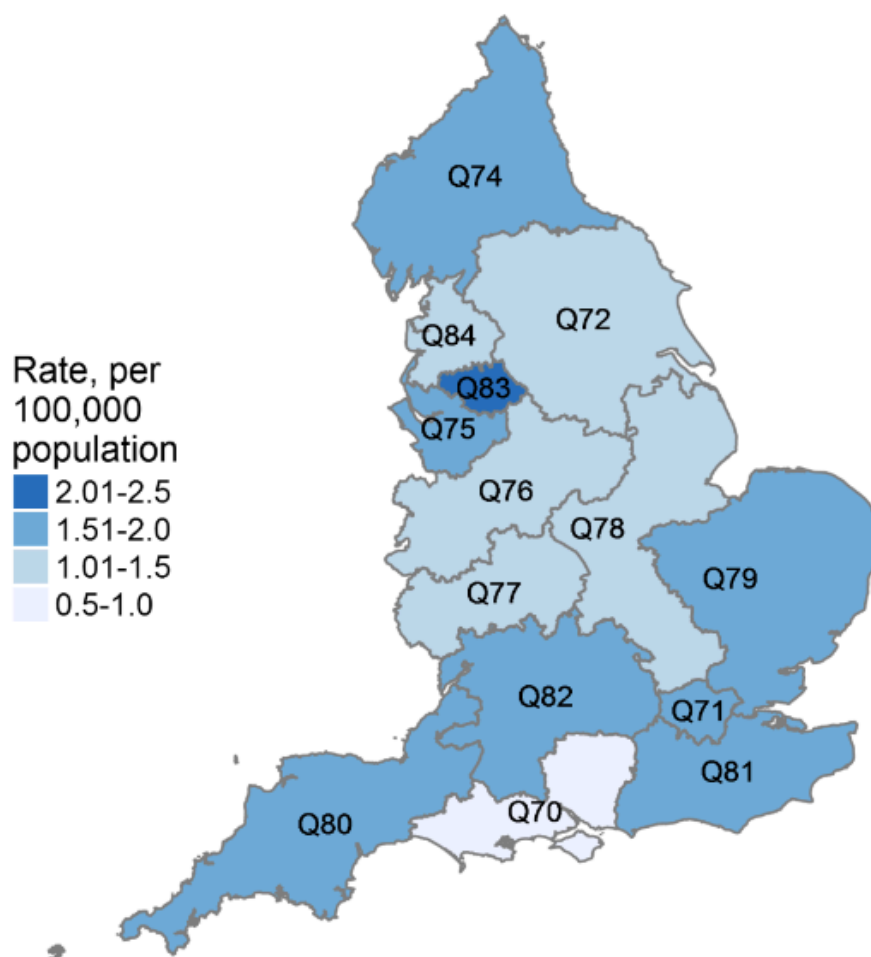
Figure 1.2. Geographic distribution of MRSA rates per 100,000 population. 2016/2017 MRSA rates are used in combination with 2015/2016 population rates as later data was not available. Reproduced from: Public Health England. Annual Epidemiological Commentary. 2016-2017.[32]

Surveillance data are also collected on whether the bacteraemia began in, or was ascribed to, the community or hospital. These data suggest that despite relatively consistent rates of MRSA bacteraemia overall on a national level, there are changes in the finer level epidemiology. Most recently, it has become apparent that a greater proportion of cases are being acquired outside of hospital Trusts, such as in community facilities managed by clinical commissioning groups (CCGs) (Table 1.1). There are a number of potential explanations for this, including changes in the circulating lineages, increasing amounts and complexity of care being delivered outside of hospitals and the focus of MRSA control initiatives in the hospital setting. Indeed, it has been proposed that the changes may be due to patients being discharged earlier and therefore symptoms of infection acquired in hospital developing only when patients are back in the community.[32]

Table 1.1. MRSA counts and rates by Post Infection Review assignment, England, from 2013/2014 until 2016/2017. Reproduced from: PHE, Annual Epidemiological Commentary, 2016/2017.

Financial year	Financial year population estimate	CCG-assigned		Total bed days	Trust-assigned		Third-party	
		Cases	Rate *		Cases	Rate**	Cases	Rate*
2013/14	53,976,973	450	0.8	34,327,781	412	1.2	NA***	NA***
2014/15	54,432,437	370	0.7	34,797,208	315	0.9	115	0.2
2015/16	54,899,107	280	0.5	34,584,084	299	0.9	244	0.4
2016/17	54,899,107	232	0.4	35,009,888	315	0.9	276	0.5

* Rates for CCG- and Third-party-assigned cases are given per 100,000 population.

** Rates for trust-assigned cases are given per 100,000 bed days.

*** Third-party assignment only commenced in 2014/15, therefore data prior to that financial year do not exist.

1.3. Identification of MRSA

1.3.1. Laboratory Identification of MRSA

Typically, laboratory identification of an antibiotic resistant bacterium would require culture and identification of the organism followed by antibiotic susceptibility testing. Whilst this is still applicable for clinical samples, the aim to identify patients with MRSA promptly upon admission to hospital, and the volume of samples obtained through universal screening, means that there has been substantial demand for identification methods which minimise processing time from screening swab to report. In England, the standard protocol issued by Public Health England details the culture of a screening swab directly onto chromogenic selective MRSA medium with or without prior broth enrichment. Incubation requires 18-48 hours, with cultures being read daily.[40]

1.3.2. Typing of MRSA

Typing confers the ability to distinguish between strains of MRSA. In order to distinguish between MRSA isolates involved in an outbreak, a high level of discriminatory ability is required when compared to typing isolates to determine their evolutionary relationships or phylogeography. There are a number of phenotypic and genotypic methods which have been used to type MRSA. Phenotypic methods include phage typing, serotyping and antibiotic susceptibility profiles. Molecular typing methods include Multi-locus Sequence Typing (MLST), pulsed field gel electrophoresis, Multiple-Loci Variable Number Tandem Repeat Analysis, staphylococcal protein A (*spa)* typing, and whole-genome sequencing (WGS). Currently, antibiotic susceptibility profiles (ASP) tend to be the most commonplace surrogate typing method used to compare isolates in clinical laboratories, whilst molecular methods such as MLST, SCC*mec* and *spa* typing are routinely used by

referral laboratories such as the Staphylococcal Reference Laboratory at Public Health England, Colindale.[41]

1.3.3. Multi-locus Sequence Typing

First developed by Maiden *et al.* to type *Neisseria meningitidis,*[42] MLST is a well-established typing method which utilises the variation within seven, highly-conserved housekeeping genes to determine bacterial relatedness. The essential metabolic functions of these genes means that changes to nucleotide sequences are slow and therefore provide insight into bacterial evolution over large timeframes.[43] Once the nucleotide sequence of each locus has been determined, the data are entered into an online centralised database (https://pubmlst.org/) and an allelic profile allocated. Seven numbers are allocated, one for each locus, forming an allelic profile, and a corresponding 'sequence type' (ST) is allocated. STs are then grouped into clonal complexes (CC) based on relatedness.[44] The population structure of MRSA is highly clonal, with low levels of homologous recombination, and therefore MLST relatively accurately reflects the population structure.[45] It is also reproducible between laboratories, and easy to collate and compare via online databases. The requirement to sequence seven genes acts as a barrier to widespread use since this is time-consuming, expensive and requires specialist equipment.[25]

1.3.4. *spa* typing

spa typing is a single-locus genotyping method that is routinely used for typing MRSA. It involves determining the variation in nucleotide sequence and number and type of repeats within the highly variable X-region of the staphylococcal Protein A gene.[46] After

determining the sequence using PCR, the nucleotide sequence is entered into an online centralised database in order to determine the *spa* type (e.g. https://www.spaserver.ridom.de).[47] Resolution is comparable to multi-locus methods such as MLST, but being a single-locus method, *spa* typing does not indicate genetic relatedness. It is, however, cheaper than MLST and it is possible to apply a clustering algorithm, known as Based Upon Repeat Pattern (BURP), to group related *spa* types together into clusters.[48]

1.3.5. SCC*mec* typing

PCR typing of the SCC*mec* type can provide useful supplementary data to MLST and/or *spa* typing of MRSA (e.g. ST8-t008-MRSA-IV). The SCC*mec* type is often used to provide additional information on the isolate, based on the less stable MGEs within the genome. Typically, the larger SCC*mec* types (I, II and III) are more abundant within lineages regarded as hospital associated, whereas the smaller SCC*mec* types IV and V are often found within those more commonly regarded as community clones, although this is less informative now that community clones are increasingly common in hospitals, and vice versa.[49] [50]

1.3.6. Antibiotic Susceptibility Profiles

Antibiotic susceptibility profiles or antibiograms are routinely used by infection control practitioners to identify potentially related isolates. Once a clinical sample enters the laboratory, culture is set up and if a potential pathogen is isolated, identification and antibiotic susceptibility testing is carried out. To identify the organism, observation of colonial morphology, microscopy and biochemical tests are undertaken. Traditionally, antibiotic susceptibility has been determined using disk diffusion testing. Having isolated

a pure culture of the bacterium, a solution with a specific turbidity (typically 0.5 McFarland) of the organism is spread across an agar plate (often Müller-Hinton agar with or without supplementation). Paper disks containing pre-defined concentrations of antibiotic agents are placed on the inoculated agar, and the plates incubated in certain conditions for defined time periods, normally between 16 and 20 hours.[51] Measurement of the size of the zone of inhibition (no growth) around the disk are compared to predefined zone sizes to determine whether the bacterium is susceptible or resistant to each antibiotic.[52] The most rapid phenotypic susceptibility testing is known as direct susceptibility testing (DST), where an unstandardized volume of inoculum is plated directly from the clinical sample to susceptibility testing, in some cases reducing time to provisional results from about 36 hours to 12 hours. However, for formal antibiotic susceptibility with a standardised volume of inoculum, the need to culture the organism during processing means that at least 36 hours of processing is required in the laboratory.

1.4. Community-associated MRSA

1.4.1. Emergence of CA-MRSA

In the 1990s, reports of MRSA infections in individuals lacking 'typical' MRSA risk factors (section 1.2.1) began to emerge. Using epidemiological criteria, the absence of prior healthcare contact meant that these infections were defined as community-associated MRSA (CA-MRSA). Although first described in Australia,[53] the United States (US) has experienced the most extensive spread of CA-MRSA.[50,54] Early investigations showed that rather than representing 'overspill' of hospital-associated strains into the community, these community strains were genomically distinguishable from the hospital strains.[55] Indeed, the early strains which were defined as CA-MRSA tended to carry the smaller SCC*mec*

types IV and V, and Panton-Valentin Leucocidin (PVL), a virulence factor associated with SSTI and necrotising pneumonia. CA-MRSA strains also carried fewer antibiotic resistance genes than hospital associated lineages.[55] The combination of epidemiology and genomics within the definition of CA-MRSA has proven challenging, with CA- lineages spreading within hospitals, and no stable genomic marker or antibiotic susceptibility pattern representing community association.

1.4.2. USA300 MRSA

The CA-MRSA pulsotype USA300 has proven particularly troublesome in the US. It was first reported in 1999 and this was followed by widespread dissemination and an epidemic of MRSA infection in otherwise healthy people.[56,57] A community associated clone, by 2004, this lineage was responsible for up to 97% of skin and soft tissue infection (SSTI) presenting to US emergency departments (ED).[50] USA300 also causes invasive disease, such as pneumonia and osteomyelitis, and over time USA300 has become endemic in US hospitals where it causes hospital-associated infections including bacteraemia.[58,59] USA300 is readily transmitted within households, which act as long-term reservoirs associated with repeated episodes of infection and onward transmission.[60,61] Antimicrobial resistance to macrolides and fluoroquinolones is common.[50] An epidemic of CA-MRSA infection caused by a clone that is closely related to USA300 but which arose independently (USA300 Latin-American variant, USA300-LV) has also been identified in South America.[62]

International travel is an important contributor to the inter-continental spread of infectious diseases,[63] and the spread of USA300 and USA300-LV have been documented

globally.[56,64] However, in the UK, the vast majority of MRSA infections continue to be caused by the dominant hospital-associated lineage EMRSA-15 (ST22).[65] USA300 is considered to be of low prevalence in continental Europe, presenting primarily as sporadic cases and discrete, small outbreaks.[66-72] Similarly, sporadic cases and a single hospital outbreak caused by USA300 have been described in the UK,[73-76] but the prevalence of USA300 carriage and infection is unknown.

1.5. MRSA Prevention and Control

1.5.1. Infection Surveillance

Surveillance is defined as 'the systematic and continuous collection, analysis and interpretation of data, closely integrated with the timely and coherent dissemination of results to those who have the right to know so that action can be taken.'[77] More simply, this can be considered as 'information for action.' Infectious disease surveillance systems are typically divided into active and passive surveillance methods. Active surveillance refers to the active solicitation of data, whereas passive surveillance refers to those systems that utilise routinely collected information to monitor infection, with the responsibility lying with the laboratory or health provider.[78] Methods used in both groups have advantages and limitations, depending on the infection, population and aim of surveillance. Importantly, active surveillance is more labour-intensive and expensive than passive surveillance. Passive is more sustainable over longer time periods, but is dependent on the quality of data recorded for an alternative purpose, and therefore vulnerable to under-reporting.

Syndromic surveillance refers to the monitoring of rates of infection based on the occurrence of a group of signs and symptoms (syndrome), which may be caused by that

infection. In the UK, the general practitioner is the first point of contact for most patients and consequently, the majority of infections will be managed by these teams in the community. The short duration of most illnesses and the time frame associated with submission and processing of samples in the laboratory mean that until point-of-contact testing is commonplace, general practitioners rely on 'syndromic diagnosis' alongside a judgement of risk of patient deterioration, in order to determine their management strategy.[79] Monitoring infection rates can be undertaken by extracting data on symptoms, signs, clinical diagnosis and treatments from clinical coding systems. Syndromic surveillance therefore monitors rates of infections where no organism has been identified, providing data trends where otherwise no information would be available. In addition, in the case of an outbreak of a novel pathogen, such as the emergence of Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV), syndromic surveillance is a useful proxy to monitor spread.[80]

Sentinel surveillance is also a frequently used mechanism for infectious disease surveillance. Here, selected facilities with the appropriate staff and facilities for diagnostics are used to provide higher quality data than that obtained through passive surveillance.[81] Sentinels are selected according to the population they serve and access to testing. In the UK, sentinel surveillance forms an important part of the national influenza surveillance programme. Here, during a pre-defined time period corresponding with seasonal influenza activity, a selection of general practitioners submit both syndromic data (consultations for influenza-like illness) and also nose and throat swabs from patients presenting with this syndrome.[82] Such a sentinel system can report trends, which can be generalised to a greater population and therefore acts as a useful alternative to passive surveillance. Whichever

method of surveillance is implemented, it is important that data are collected systematically.[83]

1.5.2. National Surveillance of Infectious Disease in England

There are a number of surveillance mechanisms for infectious diseases in England. These consist of mandatory, enhanced and voluntary systems,[84] and are a combination of passive, active, sentinel and syndromic systems. Public Health England is a government organisation which is responsible for the co-ordination of national infectious disease surveillance. Over the past 20 years, there have been various systems implemented to systematically capture laboratory data for surveillance purposes. The first national electronic system was known as COSURV/LabBase2. This system involved the electronic transfer of data from local laboratory information management systems (LIMS) to the regional PHE centre. Here, it would be processed before electronic transfer from the regional centre to LabBase2,[84] where analysis could be undertaken centrally prior to feedback to the regional/local level. An additional system, AMSURV, was introduced in 2009 to electronically capture antibiotic susceptibility profile data to enhance national AMR surveillance. In 2014, both systems were replaced by the Second Generation Surveillance System (SGSS). This currently receives voluntary submissions of laboratory data from 98% of English microbiology laboratories via electronic transfer.[85]

1.5.3. National Surveillance of MRSA Bacteraemia in England

The sharp rise in MRSA infections in the 1990s was seen in laboratory surveillance data that were submitted voluntarily to the national database, LabBase2. In response to the increase, the UK Government (Department of Health) introduced a number of control

measures, including mandatory surveillance of MRSA bacteraemia from 2001.[32] Indeed, for over fifteen years, English hospital Trusts have been required to submit data to Public Health England (PHE), as part of a national surveillance system to monitor rates of MRSA bacteraemia.[38] Initially, these data consisted of aggregated counts of MRSA bacteraemia cases, but in 2005 enhanced epidemiological data were requested to provide more information about individual cases. When studied retrospectively, this mandatory surveillance data now provides reliable information on the trends of MRSA bacteraemia in England, with strong epidemiological data on each case. It also has an ascertainment rate that is 20% higher than the comparable voluntary surveillance system.[86]

1.5.4. Post Infection Review of MRSA Bacteraemia Cases

Since April 2013, the NHS in England has adopted a 'zero-tolerance' approach to preventable MRSA bacteraemia.[87] To this end, at the time of writing, PHE co-ordinate a 'Post Infection Review' (PIR) process for each case. The aim of this is to identify the organisation where the learning from occurrences would be best placed, be it the acute Trust, CCG, or a third party in the case of complex, unassignable cases.[88] This is a multi-disciplinary and cross-organisation process which is equally applied to providers in the community and hospitals. A tool-kit is provided by PHE, which is shown in Figure 1.3. After summarising the patient journey through healthcare services, infection control, devices, and antimicrobial therapy, an action plan for learning outcomes is produced and the case apportioned to one of the organisations. As of April 2018, PIRs will no longer be required for every case of MRSA bacteraemia, but instead, only those assigned to organisations with the highest rates of infection. These organisations will be determined by PHE each year based on past performance.[89]

```
A.  Case details and timeline of healthcare contact
B.  Screening procedures applied
C.  Medical devices in situ
D.  Appropriateness of antimicrobial therapies
E.  Skin integrity
F.  Risk factors for transmission
G.  Hand hygiene
H.  Other factors
I.  Organisational issues
J.  Governance issues
```

```
Summary of learning outcomes
K.  Statement of good practice
```

```
K.  Assignment of case to organisation
```

Figure 1.3. Summary of post infection review toolkit. Adapted from: NHS England. Guidance on the reporting and monitoring arrangements and post infection review process for MRSA bloodstream infections from April 2014 (version 2).[88]

1.5.5. Local Prevention and Control of MRSA

Microbiology services in England are tiered from national to regional to local level. Since the 'Independent review of NHS Pathology Services in England' reports were published in 2006 and 2008, there has been widespread consolidation of microbiology laboratory services.[90] This means that laboratories may process samples for a wide geographical area, including hospitals, community healthcare facilities and general practices. However, surveillance and infection control within these distinct areas are often quite different.

1.5.5.1. Infection Control of MRSA in Hospitals

Infection control in hospital aims to minimise the risk of acquiring an infection when in contact with the healthcare environment.[91] For MRSA, methods focus on limiting direct transmission. Hand hygiene is a general infection prevention method which is vitally important in limiting MRSA spread, as MRSA has been shown to be transmitted on the hands of healthcare workers.[92] Specific infection control interventions vary dependent on local guidelines, but preventative measures include screening for MRSA carriage and for those identified as carrying MRSA, decolonisation, isolation and contact precautions are implemented.

At the institutional level, hospitals adopt a number of mechanisms for determining the presence and overall prevalence of pathogens within their in-patient populations. The rationale behind identifying MRSA positive individuals through screening is that isolation and contact precautions can be applied for these cases. Such interventions have been shown to interrupt the transmission of MRSA in hospital settings and reduce the risk that the individual will develop MRSA infection.[93-98] From 2008 until 2014, universal MRSA screening of hospital in-patients (including both elective and emergency admissions) was undertaken. This consisted of multi-site screening of the patient (typically three sites of the nose, throat, axilla, groin) either at the start of their admission or at a pre-admission clinic. More recently, Robotham and colleagues, 2016, undertook an effectiveness and cost-effectiveness analysis of MRSA screening in the UK NHS and showed that universal screening was unlikely to be cost-effective at concurrent NHS willingness-to-pay thresholds.[99] Therefore, the recommendation of universal screening was revised by the Department of Health MRSA Screening Implementation Group, and limited to targeted screening of high-risk populations only.[100] This consists of those previously infected with

or colonised by MRSA, and patients admitted to high-risk units. Currently, those areas deemed high-risk are summarised in Table 1.2.

Table 1.2. High-risk areas where MRSA screening is recommended by the Department of Health MRSA Screening Implementation Group.[100]

Defined High-risk Area	Areas Subject to Local Risk Assessment
Vascular	Specialised services (e.g. transplant)
Renal/Dialysis	Units with history of high endemicity
Neurosurgery	Units with patients at risk of poor outcome
Cardiothoracic surgery	
Haematology/Oncology/BMT	
Trauma and orthopaedics	
Intensive or high care units	

Isolation and contact precautions typically consist of an individual room, gloves and gowns for staff entering the room and for patient contact, and being positioned after other patients on operation or procedure lists to reduce the risk of onwards transmission. The regimen for decolonisation therapy varies depending on the institution, but the CUH regimen is outlined in Figure 1.4. Substantial cost is incurred through implementing such regimens. The most recent, comprehensive cost analysis was undertaken by Roth *et al.*, 2018, and showed through a prospective cost analysis in a Swiss hospital that contact precautions alone cost approximately US$158.90 per patient per day.[101] In addition, Cairns et al., 2014,

determined that screening based on clinical risk assessment may lead to similar rates of detection with greater cost effectiveness than universal screening.[102] In practice, isolation and contact precautions are continued until MRSA colonisation has been shown to be no longer present.[100] Depending on the hospital, patients in England tend to be considered MRSA negative if three consecutive screens are negative.[95,103]

| **Mupirocin 2% nasal ointment**
Applied to both nostrils three times per day for five days | **Rescreen** 48hours after treatment has ended and then at weekly intervals for 3 weeks. |
| **Octenisan wash concentrate**
Applied neat to moistened skin, left for one minute and then washed off
Shampoo hair with Octenisan day 1 and 4 | |
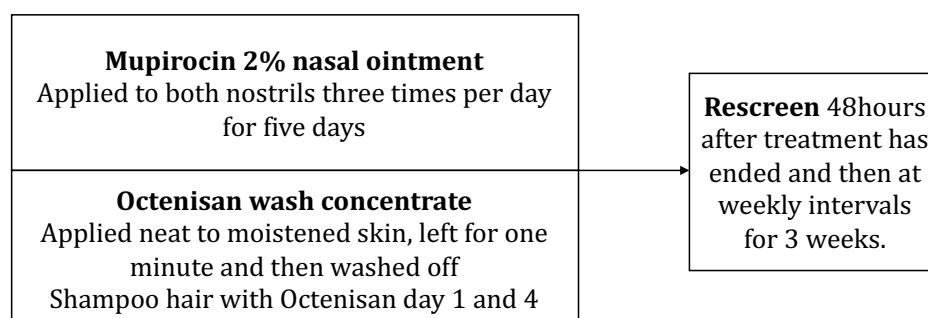
Figure 1.4. CUH regimen for MRSA decolonisation

In each hospital, the infection control team will receive daily updates of the cases of MRSA, often through an automated IT system. The infection control team (ICT) will then review the patients, advise on decolonisation therapy and determine appropriate infection control measures to minimise the risk of transmission. If there is more than one case of MRSA with epidemiological links, the ICT may investigate further regarding a potential outbreak (Section 1.5.5.3).

1.5.5.2. Prevention and Control of MRSA in the Community

In contrast to hospitals, there is greater variation in the structure of infection control services in the community. The responsibilities of organisations providing healthcare in the community, however, are well-defined. Like hospitals, general practices and similar

healthcare facilities are required to meet infection control standards outlined in the Health and Social Care Act 2008: Code of Practice on the prevention and control of infections and related guidance.[104,105] The Care Quality Commission then assesses a provider against 10 pre-defined criteria to determine compliance to the code of practice (Table 1.3).

Table 1.3. Table summarising the 'Code of Practice' for providers of healthcare on the prevention of infections under the Health and Social Care Act 2008. A registered provider is judged on how it complies with each regulation related to infection prevention outlined below. *Reproduced* from: The Health and Social Care Act 2008 Code of Practice on the prevention and control of infections.

Compliance criterion	What the registered provider will need to demonstrate
1	Systems to manage and monitor the prevention and control of infection. These systems use risk assessments and consider the susceptibility of service users and any risks that their environment and other users may pose to them.
2	Provide and maintain a clean and appropriate environment in managed premises that facilitates the prevention and control of infections.
3	Ensure appropriate antimicrobial use to optimise patient outcomes and to reduce the risk of adverse events and antimicrobial resistance.
4	Provide suitable accurate information on infections to service users, their visitors and any person concerned with providing further support or nursing/ medical care in a timely fashion.
5	Ensure prompt identification of people who have or are at risk of developing an infection so that they receive timely and appropriate treatment to reduce the risk of transmitting infection to other people.
6	Systems to ensure that all care workers (including contractors and volunteers) are aware of and discharge their responsibilities in the process of preventing and controlling infection.
7	Provide or secure adequate isolation facilities.
8	Secure adequate access to laboratory support as appropriate.
9	Have and adhere to policies, designed for the individual's care and provider organisations that will help to prevent and control infections.
10	Providers have a system in place to manage the occupational health needs and obligations of staff in relation to infection.

What is of great contrast in the community versus hospital setting is the reduced infection control manpower and expertise. Whereas each hospital will employ specialist infection control doctors and nurses, in the community setting an 'infection control lead' will hold

responsibility for the compliances detailed in Table 1.3. This individual is likely to be a general practitioner, practice nurse or practice manager who has undertaken specific infection control training, rather than an infection specialist.

There are, however, overarching teams which offer support to individual practices. Using the Cambridgeshire region as an example, Cambridgeshire and Peterborough CCG employs two infection control nurses to provide support and specialist advice to facilities commissioned by the CCG in the region (e.g. general practices and nursing homes).[106] In addition, those infections deemed to pose a risk to public health are managed with the PHE East of England Health Protection Team, which consists of health protection practitioners and consultants in communicable disease control. Some NHS hospital microbiology teams also provide infection control advice to practitioners in the community, although the extent of this may vary dependent on the contract of service between the laboratory and practitioners. In Cambridgeshire, however, general practices are served by the PHE laboratory at Cambridge University Hospitals NHS Foundation Trust. At the time of writing, there are currently eight PHE laboratories across England. PHE laboratories aim to deliver "microbiology support for the investigation, management and control of infection and outbreaks of communicable disease."[107] Each PHE laboratory serves a defined geographical region and employs a regional public health microbiologist to oversee the delivery of public health microbiology in the region. Where NHS laboratories are contracted to provide community services (this is not the case in the Cambridgeshire area), the PHE laboratory will act as a 'back-up' if particular public health functions cannot be delivered.[107]

MRSA is not considered a 'notifiable' organism of concern to public health, and cases or outbreaks in the community do not automatically require input from public health teams. Therefore, in Cambridgeshire, the author's experience has shown that cases tend to be managed by community practitioners with advice from the CCG infection control nurses, and advice from hospital/PHE microbiologists if sought.

Typically, general practitioners have only been required to screen individual patients for MRSA in exceptional circumstances, such as a patient having surgery in a hospital within a different region.[103] If an individual is positive on screening but asymptomatic and not due to attend hospital, then decolonisation is not required. If MRSA is suspected clinically and sampled, or detected unexpectedly on an infected clinical sample, then decolonisation is advocated. If no specific policy is present, general practitioners may be advised to follow local hospital protocols regarding decolonisation therapy (e.g. Figure 1.4).[103]

Increasing amounts of complex care and procedures are being moved from hospitals to the community, as outlined in the "NHS Five Year Forward View," 2014.[108] Given this, relevant infection control measures which are implemented in secondary care may need to be transferred to primary care if a patient is MRSA positive. For example, in secondary care, patients with MRSA are placed at the end of procedure lists to reduce the chance of cross-contamination. In primary care, this is more difficult to administer, for example, as appointments may be allocated based on both patient and provider availability. In addition, not all healthcare in the community is delivered in a clinic setting, for example some professionals routinely undertake healthcare tasks in patient's homes. Consensus opinion would suggest that infection control practices would be beneficial in preventing the spread of infection in the community setting, but there are few studies in this area and so evidence

is lacking. It has, however, been shown that some practices which have become highly embedded within hospital practice (e.g. hand hygiene) are not always implemented in general practice.[109-111] Despite this, the professional guidelines for healthcare professionals and regulations for registered providers of healthcare, as outlined above, do stress the importance of infection control to prevent infection transmission.

1.5.5.3. Identification and Definition of an MRSA Outbreak

Both in the community and within hospitals, an outbreak is typically defined as 'a localised increase in the incidence of disease.'[77] In practice, the term cluster and outbreak are used interchangeably; indeed, it can be argued that the term used may be selected based on how 'serious' terminology needs to sound, with an outbreak sounding more substantial than a cluster. Detecting outbreaks can be labour intensive. Historically, outbreaks have tended to be discovered if they involve an unusual (e.g. antibiotic resistant) organism, an unusual site of infection, or cause a large increase in infections. More systematic analysis of surveillance data can lead to more objective detection of outbreaks. Both methods have substantial limitations, ranging from a reliance on human judgement ("gut feeling"), to the inflexibility of fixed rules which may miss true outbreaks.

In a typical diagnostic laboratory, the pattern of MRSA susceptibility to different antibiotic agents provides the only readily available additional information regarding strain type. Therefore, the ASP is often used when considering whether isolates that are associated in time or space are related. If the ASPs and basic epidemiology align, this may indicate transmission or an outbreak. Samples may then be sent to the PHE Staphylococcal Reference Laboratory for further typing to investigate the relatedness of the isolate at a higher resolution. The logistics of sending samples to another site for investigation incurs

a further time delay of a number of days. Consequently, the initial evaluation of the need for infection control intervention will need to be made using ASP data only. In 1999, Grundmann *et al.* attempted to determine the potential of laboratory services in determining the cross-transmission of *S. aureus*. They found that the majority of isolates were from two genotypes, as determined by PCR fingerprints. Despite this, most of the patients were shown to be epidemiologically unrelated. ASPs have also been shown to act as 'triggers' for further investigation using genomics in possible outbreaks which have then been halted.[112,113] With the increased use of WGS in research, a number of studies have undertaken small investigations into ASPs versus WGS for detecting MRSA transmission. Harris *et al.*, 2013, investigated an outbreak using WGS and identified that two isolates were incorrectly excluded due to variation in ASP from the outbreak strain.[114] Similarly, Azarian *et al.*, 2015, showed through study of MRSA dynamics on a neonatal intensive care unit that despite minimal variation in ASP, 10 of the 17 USA300 isolated were unrelated.[115] Gordon *et al.*, 2018, concluded in their study that ASPs could lead to isolates being falsely excluded from outbreak investigations.[116]

Despite being a cornerstone of infection control practice in both triggering and guiding outbreak investigations, at the time of writing there remains no systematic analysis of the reliability of ASPs and epidemiology in detecting MRSA transmission.

1.6. Whole-Genome Sequencing

1.6.1. History of WGS

1.6.1.1. First Generation 'Sanger Sequencing'

The era of whole-genome sequencing began in the 1970s when the first complete DNA genome, that of phiX174 bacteriophage, was sequenced by Fred Sanger and his team.[117] Known as Sanger sequencing, in this process nucleotides are incorporated into 450-500bp fragments complementary to the template DNA strand.[118] Each strand is terminated by the integration of a modified nucleotide, known as ddNTP, which also acts as a label for the base *in situ* to be read, typically via gel or capillary electrophoresis.[119] The high level of accuracy achieved through Sanger sequencing is reflected by high time and cost commitments.

1.6.1.2. Second Generation Sequencing

It is with these limitations in mind that 'second generation sequencing' technologies were developed. Also known as 'next generation sequencing', a number of methods have been developed. The technologies used and discussed in this work have been developed by Illumina, (San Diego, USA), formerly Solexa, (Cambridge, UK), which is currently the dominant provider of second generation sequencing technologies.[120]

Each time that a sequencing experiment is undertaken and sequencing data is generated, it is known as a 'run'. Small genome sequences generated through sequencing are known as 'reads.' These are assembled by piecing together based on overlaps in the short sequences (*de novo* assembly) or by comparing to a predetermined, related reference genome for

which the sequence is already known (mapping). Because the process is error-prone, each position is sequenced many times in many reads. The overlap in these reads is known as the 'coverage'.

Second generation sequencing methods prioritise the parallel processing of samples to reduce the reaction time and cost, which limited first generation technologies. Illumina sequencing works through 'sequencing by synthesis'. First, native DNA is prepared by the creation of a library, the production of DNA fragments of 300-500 base pairs in length. Adaptors are attached to these DNA fragments to aid ligation to the flow cell and act as primers for PCR amplification. Fluorescently tagged nucleotides are used to determine which base is incorporated as the synthesis proceeds, and the sequence is determined. Paired-end sequencing is undertaken to provide positional information about the genome. In order to achieve increased sample throughput, multiplexing has become an important feature of the Illumina platform. Tags are added to the adaptors in the sequencing libraries. These tags are recorded and then are sequenced together and separated at the analysis stage. Despite this method being higher-throughput, only 100-150 base pairs are of sufficient quality and so Illumina technology is limited to generating short read data.

1.6.1.2. Third Generation Sequencing

The current limitations of second generation sequencing methods such as that provided by the Illumina platform are the time required per run and the length of high quality sequence that is achievable. 'Third generation sequencing' methods have been designed with the aim of addressing these.

Of particular interest to public health is the MinION (Oxford Nanopore). This is a highly portable technology which can be used with minimal specialist equipment, and generates DNA sequences in real-time using nanopore technology. To summarise, DNA strands associate with a pore which generates an electrical signal, detected by a sensor. This technology is capable of generating much longer sequences than NGS technologies, but currently does not produce as high quality data. It has, however, been shown by Quick *et al.*, 2015, that this technology could determine if Salmonella samples were part of a hospital outbreak within two hours.[121] Since then, studies have shown the application to identification of transmission of carbapenemase-producing *Klebsiella pneumoniae* and a resistant lineage of *Serratia marcescens*.[122,123] Once this technology and the subsequent analysis is optimised, the combination of long read, real-time data, compatability and portability will make this of particular interest in public health microbiology.

1.6.2. WGS and MRSA

1.6.2.1. Insights into Evolution and Population Structure of MRSA Using WGS

In the last decade, a number of seminal studies have examined the spread and population structure of MRSA across the globe using whole-genome sequencing. In 2010, Harris and colleagues were the first to demonstrate the use of WGS in studying the evolution and population structure of ST239 MRSA, a multiple antibiotic resistant lineage which at the time was responsible for over 90% of HA-MRSA in Asia. They sequenced 43 isolates from a global collection of ST239 between 1982 and 2003, and 20 isolates from a Thai hospital. The subsequent phylogenetic analyses demonstrated strong evidence for both intercontinental spread and hospital transmission of ST239.[124] McAdam *et al.* then studied the evolution of the endemic UK lineage by sequencing 60 E-MRSA16 lineage and 27

contextual CC30 isolates. Analyses dated the emergence of E-MRSA16 to approximately 35 years previously and suggested that this lineage had spread from large urban centres such as London and then through other regions.[125] Also relevant to the UK, Holden *et al.*, 2013, sequenced 193 MRSA ST22 isolates from around the world to explore the evolution and emergence of EMRSA-15, the dominant clone in the UK. They determined that the current lineage of EMRSA-15 had emerged from an ancestor which had spread across the UK in the 1980s. The current lineage was determined to have emerged in the Midlands in the 1980s, and the acquisition of fluoroquinolone resistance may have been pivotal in this dissemination.[126]

1.6.2.2. Local MRSA Outbreak Investigation Using WGS

WGS offers such a high level of resolution that it is possible to distinguish between isolates with one base difference. Because of this, it has offered substantial advances in the investigation and management of potential infection outbreaks in hospitals and the community. In one of the earliest studies demonstrating this application, Köser *et al.*, 2012, retrospectively sequenced isolates from an outbreak on a neonatal intensive care unit. They studied isolates from seven case and control babies and phylogenetic analysis revealed that it was possible to clearly distinguish the cluster of outbreak strains from the control isolates. They also determined a previously unknown transmission event between two control patients.[112]

Building on from this, Harris *et al.*, 2013, used WGS in real time during a suspected outbreak of MRSA linked to a special care baby unit (SCBU).[114] This investigation identified 26 related cases of MRSA infections and/or carriage, and showed that

transmission occurred within the SCBU, between mothers on a postnatal ward, and in the community up until the end of 2011. Epidemiological investigation was complicated by variation in the antibiotic susceptibility profile of the isolates. The outbreak strain was a novel multilocus sequence type, ST2371, related to the dominant hospital-associated lineage in the UK (ST22, EMRSA-15) but Panton-Valentine leucocidin (PVL)-positive. A case which occurred 64 days after the previous cases triggered healthcare worker screening and led to the detection of a worker carrying the outbreak lineage. Decolonisation therapy was administered and no further cases identified. Standard infection control measures had failed to characterise isolates involved in the outbreak due to the variation in antibiogram, which misled the ICT regarding isolate relatedness. This also demonstrates the ability of WGS to supersede reference laboratory typing as a method of confirmation of isolate relatedness, benefiting from discriminatory ability and reduced logistical testing constraints.

Similar studies have now been undertaken around the world, demonstrating the practicality of using WGS to study MRSA outbreaks in hospitals when traditional infection control methods have suggested an outbreak may have occured.[113,116,127,128] Whilst the value of improved detection and therefore early intervention of outbreaks is easily demonstrated, Török *et al.*, 2014, also demonstrated that being able to exclude an outbreak using WGS is also beneficial as it allows a step down of control measures, avoiding unnecessary cost.[129]

1.6.3. Surveillance of MRSA in England using WGS

In the Chief Medical Officer report, 2011, the significant opportunity offered through novel technologies such as whole-genome sequencing and point of care testing to undertake

'more rapid and more discriminating surveillance,' was discussed.[130] Whilst the surveillance strategies discussed previously (Section 1.5.2) are integrated within laboratory systems and operate continuously and sustainably within Public Health England, other organisations also undertake national surveillance programmes for MRSA. The British Society for Antimicrobial Chemotherapy (BSAC) have been undertaking surveillance for antibiotic resistance for over 15 years. A selection of British laboratories are invited to submit 20 *S. aureus* isolates to BSAC for characterisation each year.[131] Furthermore, the European Centre for Disease Control (ECDC)-funded European Antimicrobial Resistance Surveillance Network (EARS-Net) is a Europe-wide surveillance programme to which selected sentinel laboratories submit the first five invasive MRSA isolates processed each month.[132]

Isolates from these collections have been used in research projects to explore the population structure of MRSA across Britain and Europe. Reuter *et al.*, 2016, showed that the WGS of 1013 MRSA isolates collected through the BSAC surveillance programme between 2001 and 2010 could be used as a resource for investigation of outbreaks and the tracking of antibiotic resistance in England.[76] Similarly, Aanensen and colleagues sequenced the genomes of 308 *S. aureus* isolates collected from EARS-Net surveillance across Europe (123 of which were MRSA), to demonstrate the importance of representative national and international data sets in genomic infection surveillance, and also to show how bioinformatic tools are required to manage the data which is generated.[133] Both studies demonstrate the potential for isolate collections from surveillance programmes to act as genomic frameworks to monitor specific lineages and contextualise outbreaks. However, both were one-off studies undertaken and grant-funded through academic research and currently, there is no real-time genomic surveillance programme for MRSA in England.

1.6.4. Challenges to Implementation of WGS in Microbiology Services

Whilst the potential benefits and proof-of-principle can be evidenced through academic research, the implementation of new technologies into clinical and public health practice is not a quick process. WGS is now routinely used within academic institutions. However, it remains distant from most English clinical microbiology laboratories, which continue to rely on ASPs, and in the minority of centres, single gene PCR, to type isolates in-house and then send the isolate away to the Staphylococcal Reference Laboratory for further characterisation. Only in April 2017 did the PHE Staphylococcal Reference Laboratory switch to routinely undertaking whole-genome sequencing to characterise referred isolates rather than *spa* typing and PCR determination of virulence genes. Therefore, despite the rapid development of WGS, there remain a number of challenges to overcome prior to implementation of WGS into routine practice.

Work is on-going to improve WGS direct from clinical samples including directly from blood cultures.[134] Whilst WGS has become increasingly rapid and third generation technologies such as Nanopore offer further progress in achieving identification and susceptibilities promptly, when a pure culture is required, there will always be a substantial time-lag between taking a sample and achieving WGS data. Once it is possible to sequence from a clinical sample directly, point-of-care WGS becomes possible, with the associated benefits of rapid testing.

Routine WGS will generate vast amounts of data, which will require substantial infrastructure for transfer and quality control. Automated data handling and analysis tools will be required to ensure data is stored appropriately, and that non-bioinformaticians can

interpret and act upon WGS data. Vast amounts of this data will not be of use in day-to-day clinical practice, and will need to be presented appropriately.

Curated, centralised, open access databases will be required in order for locally generated data to be contextualised. This is essential for outbreak analyses. Tools to interrogate such databases are already under development.[135,136] Standardised protocols for the collection and processing of genomic data will also be needed. Further to this, it is necessary that epidemiological and genomic data are integrated within a database which is accessible by relevant individuals (e.g. clinicians and public health professionals). Whilst this does raise concerns regarding ethics and confidentiality, similar web systems, albeit without genomic data, are in routine use. HPZone, a web accessible case management database which contains sensitive data, is used daily by PHE health protection practitioners (http://hpzoneinfo.in-fact.com/).

Whilst academic institutions can fund the upfront and on-going costs for WGS through grant-based funding, PHE and NHS funded laboratories are public services and it would be untenable to invest in such expensive technology without established evidence of benefit. For slow-growing organisms such as *Mycobacterium tuberculosis*, the long process of culture means that current practice is considered sub-optimal. Determining antibiotic susceptibilities from WGS data is therefore considered strongly beneficial to both patient care and public health, and has progressed rapidly. Currently, genomic susceptibility testing is beginning to be trialled against routine clinical practice.[137] For organisms such as *S. aureus*, despite increasing evidence of the equivalence of genomic determination of resistance,[138-140] the relatively rapid turnaround times of culture and phenotypic methods

mean that the implementation of genomics into the routine laboratory is likely to take much

longer.

1.7. Thesis Aims and Objectives

The overall aim of this work is to use phenotypic, epidemiological and genomic data to investigate MRSA surveillance at the local, regional and national level by addressing the following objectives:

1) To systematically evaluate the reliability of ASPs and epidemiology in detecting MRSA transmission.

2) To investigate the prevalence of high-risk and unmonitored MRSA lineages in the Cambridgeshire region.

3) To explore the feasibility of the integration of genomic and epidemiological surveillance at a national level.

To this end, Chapter 3 determines the reliability of ASPs and epidemiology for detecting transmission of MRSA in Cambridgeshire. These methods are compared to whole-genome sequencing and epidemiology.

Chapter 4 investigates two potentially high-risk lineages, USA300 and ST2371, which remained undetected through routine practice but were identified using whole-genome sequencing.

Chapter 5 investigates how the typically nosocomial lineage of MRSA, E-MRSA15, was transmitted within the community and was not identified through routine practice despite causing severe infections.

Chapter 6 explores the feasibility of a combined epidemiological and genomic approach to

surveillance of MRSA bacteraemia at a national level.

Chapter 2. Materials and Methods

The methods described here are those that are common to several results chapters. Each results chapter also contains materials and methods where these are specific to the chapter study design and analysis.

2.1. "Cambridgeshire Prospective MRSA Study" Outline

A dataset generated previously by the research group in which the author was based has been central to this thesis (Appendix 2.1).[145] In brief, a 12-month prospective observational cohort study was undertaken between April 2012 and April 2013. This study identified all individuals with any MRSA-positive samples processed within the Public Health England Clinical Microbiology laboratory at Cambridge University Hospitals NHS Foundation Trust. This laboratory was responsible for the processing of samples from healthcare facilities within the region and during this time received samples from 75 general practices and four hospitals – Addenbrooke's hospital (a University teaching hospital), The Rosie Hospital (an obstetric and gynaecological hospital), Papworth Hospital (a cardiothoracic hospital), and Hinchingbrooke Hospital (a district general hospital).

2.1.1. Sampling

All samples received by the microbiology laboratory for processing as part of routine screens (screening samples) or direct clinical care (clinical samples) as requested by the

requesting clinician were included. All individuals with MRSA isolated at least once from any sample type were included in the study.

At the time of the study, in-line with national guidance, universal multi-site screening for MRSA carriage of individuals admitted to hospital and weekly screening of patients in critical care units was undertaken.[141,142] The CUH laboratory sampling protocol recommended that multi-site screens include sampling of the nose, throat and groin. There was no national policy for MRSA screens in the community and so submission of MRSA samples in non-hospital settings was dependent on individual practitioner clinical decision making.

2.1.2. Epidemiological Data Collection

For those patients admitted to hospital, clinical metadata and demographic information were collected from electronic or paper records. This included data on patient movement for hospital admissions in the preceding two years. For all patients, demographic and epidemiological data was collected including residential postcode.

2.1.3. Ethical Approval

The study protocol was approved by the National Research Ethics Service (ref: 11/EE/0499), the National Information Governance Board Ethics and Confidentiality Committee (ref: ECC 8-05(h)/2011), and the Cambridge University Hospitals NHS Foundation Trust Research and Development Department (ref: A092428).

2.1.4. Bacterial Culture and Identification

MRSA samples were received in the CUH microbiology laboratory and processed by biomedical scientist staff as per protocol (PHE standards for microbiology investigations).[40,143] In outline, MRSA was isolated from screening samples by direct plating onto Oxoid Brilliance[TM] MRSA chromogenic medium (Oxoid, Basingstoke, UK). Clinical samples were plated onto Columbia Blood Agar (Oxoid, Basingstoke, UK) and *S. aureus* identified through using a latex agglutination kit (Pastorex Staph Plus, Bio Rad Laboratories, Hemel Hempstead, UK). A single colony was selected for further processing. The identities of all presumptive *S. aureus* were confirmed using Matrix-Assisted Laser Desorption/Ionisation Time of Flight Mass Spectrometry (MALDI-ToF MS).

2.1.5. Antibiotic Susceptibility

Phenotypic antimicrobial susceptibility was determined via two methods, disk diffusion and using the VITEK 2 instrument, card P620 (bioMerieux, Marcy l'Etoile, France). Antimicrobial susceptibility was determined for the following antibiotic agents using disk diffusion: Flucloxacillin, ciprofloxacin, erythromycin, fucidic acid, gentamicin, tetracycline, vancomycin, mupirocin, rifampicin, neomycin, linezolid, chloramphenicol. Antimicrobial susceptibility results were interpreted using European Committee on Antimicrobial Susceptibility Testing (EUCAST) zone diameters.[52] Antimicrobial susceptibility to the following agents was determined using VITEK 2: Benzylpenicillin, cefoxitin, oxacillin, ciprofloxacin, erythromycin, chloramphenicol, daptomycin, fusidic acid, gentamicin, linezolid, mupirocin, nitrofurantoin, rifampicin, teicoplanin, tetracycline, tigecycline, trimethoprim, vancomycin, clindamycin.[144]

2.2. WGS Methodology

2.2.1. DNA Extraction and Sequencing

Bacterial genomic DNA was extracted using the QIAxtractor kit (QIAgen, Hilden, Germany), and sequencing of DNA for 2,320 samples was performed by the core sequencing team at the Wellcome Sanger Institute (WSI). Libraries were prepared and 150-bp paired end sequences determined on an Illumina HiSeq2000 (Illumina, San Diego, CA, USA), as previously described.[114] As quality control, 16 samples were filtered out at this point due to low numbers of reads (n=2) or duplicate/missing information (n=14).[145]

2.2.2. Sequence Data Assembly

Fastq files of genome sequence data were *de novo* assembled using an established assembly and improvement pipeline at the WSI.[145] This consisted of VelvetOptimiser v2.2.5 to determine the optimal kmer size and then Velvet v1.2 creating multiple assemblies.[146] SSPACE v2.0 and GapFiller v.1.11 scaffold the best N50 contigs and fill sequence gaps of more than one nucleotide respectively. As further quality control, 10 samples were filtered out as they were not MRSA.[145]

2.2.3. Multi-locus Sequence Typing

MLST sequence types were determined from the assemblies using MLST check (https://github.com/sanger-pathogens/mlst_check). The assembled genomes were compared against the MLST database for *S. aureus* (http://pubmlst.org/saureus/) to assign clonal complexes.

2.2.4. Mapping of Sequence Data

Sequence data were mapped to closely related reference genomes (Table 2.1) using SMALT (http://www.sanger.ac.uk/science/tools/smalt-0) as previously described.[114] In summary, reads were mapped if their identity matched the reference greater than 90%. Variation statistics at each base were calculated using samtools mpileup and bcftools view. Bases with uncertainty in the base call were removed. The bcftools variant quality score was required to be greater than 50 and mapping quality greater than 30. If not all reads gave the same base call, the allele frequency, as calculated by bcftools, was required to be either 0 for bases called the same as the reference, or 1 for bases called as a single nucleotide polymorphism (SNP). The majority base call was required to be present in at least 80% of reads mapping at the base, and the minimum mapping depth allowed was 4 reads, at least two of which had to map to each strand. If any of these filters were not met, the base was called as uncertain. Through quality control, 12 samples were filtered out (high number of heterogenous sites, n=8; abnormal position in the phylogeny, n=4).[145]

Table 2.1. Reference genomes for each CC

CC	Strain ID	Accession Number
22	HO 5096 0412	HE681097
30	MRSA252	BX571856
5	N315	BA000018
8	USA300	USA300 FPR3757

2.2.5. Phylogenetic Analysis

Mobile genetic elements were removed from whole-genome alignments (https://github.com/sanger-pathogens/remove_blocks_from_aln) to identify the phylogenetically informative core genome for each isolate, and SNPs used to create a mid-point rooted, maximum-likelihood phylogeny with 100 bootstrap replicates.[147] Trees were visualised using Figtree (http://tree.bio.ed.ac.uk/software/figtree/) and ITOL (http://itol.embl.de/). Pairwise genetic distances between isolates of the same CC were calculated based on the number of SNPs in the alignment of the core genome using an in-house script.

2.3. Other Methods

2.3.1. Statistical Analyses

Statistical analyses were undertaken in RStudio (version 0.99.489).

2.3.2. Figure Production

Figures were generated in RStudio (version 0.99.489), Microsoft Excel (version 14.6.1) and Adobe Illustrator CS5.

Chapter 3. Evaluation of Antibiotic Susceptibility Profiles as a Method of Detecting MRSA Transmission

3.1. Introduction

A primary goal of infection surveillance within a healthcare setting is to prevent pathogen transmission and detect outbreaks.[83] An outbreak is often practically defined as 'a localised increase in the incidence of disease.'[77] Historically, there has been a reliance on rule-based approaches combined with practitioner intuition in detecting possible outbreaks of infection.[148] However, over the past decade or so, a number of studies have explored the potential benefits of improved hospital infection surveillance through automated detection of links between patients using information technology (IT). This includes the potential benefits of reduced demands on practitioner time alongside more rapid rates of detection.[149,150]

As antimicrobial susceptibility profiles are a readily available piece of information in a clinical laboratory, they are often combined with patient movement (time and place) to detect and evaluate putative MRSA outbreaks. Further investigation and ward visits are then undertaken to monitor cases and investigate potential transmissions accordingly. The clonal nature of MRSA limits the use of traditional typing methods in distinguishing between isolates. However, the recent, rapid advancements in the understanding of bacterial molecular epidemiology via the development of WGS over the last 15 years has provided evidence of previously unrecognised, yet extensive, local, regional and global transmission pathways of MRSA.[112,114,124,145]

WGS provides the highest available resolution with which to discriminate between bacterial isolates and has proven particularly useful in detecting outbreaks at the local level, where a high degree of resolution is necessary to confirm or refute isolate involvement.[112,114,116,127,129] A number of studies using WGS to investigate MRSA outbreaks have highlighted the variation in ASPs within isolates which are genomically linked, implying limitations to using ASPs to determine transmission in practice.[114,116] However, to date, the validity of ASPs and epidemiological (time and space) approaches in identifying MRSA transmission has not been systematically evaluated on a larger scale.

In a large population-based study, WGS combined with high-quality epidemiological data uncovered a high level of hidden MRSA transmission in four Cambridgeshire hospitals and the surrounding community.[145] This study also showed that shared GP practice and ward contacts (other than the ED) were present in <1% of genetically unrelated cases.

This chapter builds on previous findings through an extended analysis of this dataset to evaluate the accuracy of conventional techniques employed in clinical practice (ASPs, combined with strong epidemiological links and *in silico spa*-typing) to detect genomically defined MRSA transmission events. This chapter aims to answer a question relevant to clinical practice in two separate but important settings. First, if there are two MRSA positive patients who have been in the same hospital ward at the same time and have the same or a single mismatch in ASP, what is the likelihood that these represent a putative transmission event? Second, if the same was to occur between a pair of individuals who share a post code, what is the likelihood that these two isolates represent a putative transmission?

3.2. Specific Materials and Methods

3.2.1. Study design

Isolates and associated patient metadata were sourced from the 12-month prospective observational cohort study (the Cambridgeshire prospective MRSA study, Appendix 2.1) between April 2012 and April 2013, as described in Chapter 2 (Materials and Methods). In brief, this identified individuals with MRSA-positive samples processed by the Clinical Microbiology and Public Health Laboratory at the Cambridge University Hospitals NHS Foundation Trust (CUH). A total of 2,282 MRSA isolates from 1,465 individuals were sequenced and met quality control criteria. For this analysis, only the first MRSA sample isolated from 1,465 individuals during the study period was analysed.

3.2.2. Epidemiological analysis

Epidemiological data (in-patient hospital stays and residential post codes) were recorded for all cases. Epidemiological links were established between each pair of MRSA-positive individuals (case-pairs) through systematic comparisons, as described previously.[145] In the first hospital analysis, direct ward contacts were defined as follows: a case-pair admitted to the same ward with overlapping dates of admission. Admissions to the emergency department were excluded, and only those admitted for one day or more were included in the hospital analysis (n=906). In the second community analysis, case-pairs with the same post code were considered epidemiologically linked. A total of 1367 patients had a registered post code and were therefore included in the community analysis.

3.2.3. Microbiologic evaluation and antibiotic susceptibility testing

MRSA was isolated from screening and clinical samples as per the methods described in Chapter 2 (Materials and Methods). Antimicrobial susceptibility was determined to commonly-used antibiotic agents (Appendix 3.1), using the VITEK 2 instrument (bioMerieux, Marcy l'Etoile, France). Systematic comparison of each isolate ASP was undertaken to determine similarities within case-pairs using two thresholds: an exact match in ASP, or a mismatch of a single antibiotic (Appendix 3.2).

3.2.4. WGS and MLST

MRSA sequence data was determined and CCs assigned using methods described in Chapter 2 (Materials and Methods). If the case-pair were from the same CC, pairwise genetic distances between isolates were calculated based on the number of SNPs in the core genome; if not, this was recorded as 'different CC.'

3.2.5. *In silico spa*-typing

Isolates were *spa*-typed using *in-silico* polymerase chain reaction (PCR) to extract the *spa* gene variable X region from assembled genomes using published primers.[151] The *spa*-type was then determined using an online *spa*-typer tool (http://spatyper.fortinbras.us/). Whether the *in silico spa*-types matched was then established for each case-pair.

3.3. Results

ASPs are commonly used in the early stages of infection control outbreak surveillance in routine clinical practice. The utility of ASPs was evaluated using a reference dataset of 1,465 MRSA isolate genomes from 1,465 patients, collected over a one-year period. The additional utility of *spa*-type, a common typing method used in determining MRSA relatedness, was also explored through determining the *spa*-type from WGS data (n=1398) and comparing this with the reference genomic and epidemiological datasets.

WGS was used within a population with strong epidemiological links as the gold standard for identifying putative transmissions. For the purpose of this work, a putative transmission was defined as a genetic distance of 50 SNPs or less and an epidemiological link (admission to the same hospital ward at the same time in the first analysis, or shared post code in the second analysis) between two patients (a 'case-pair'). Isolates were genomically diverse; MLST defined 58 STs that resided in 19 CCs but the collection was predominated by CC22 (1035 isolates, 71%) (Figure 3.1). Phenotypic antimicrobial susceptibility testing determined 132 unique ASPs for the 1,465 MRSA isolates, with two profiles accounting for 698 (48%) of isolates. (Figure 3.1). *In silico spa*-typing assigned 195 different *spa*-types to 1398 isolates (95%) for which this could be determined (Figure 3.1), with one *spa*-type (t032) predominating (n=631, 45%). Delineation of the number of ASPs for each CC and *spa*-type demonstrated a high degree of ASP diversity within isolates belonging to the same CC or *spa*-type. CC22 and CC30 contained 74 and 19 different ASPs, respectively (Table 3.1), while t032 contained 53 different ASPs (Appendix 3.3).
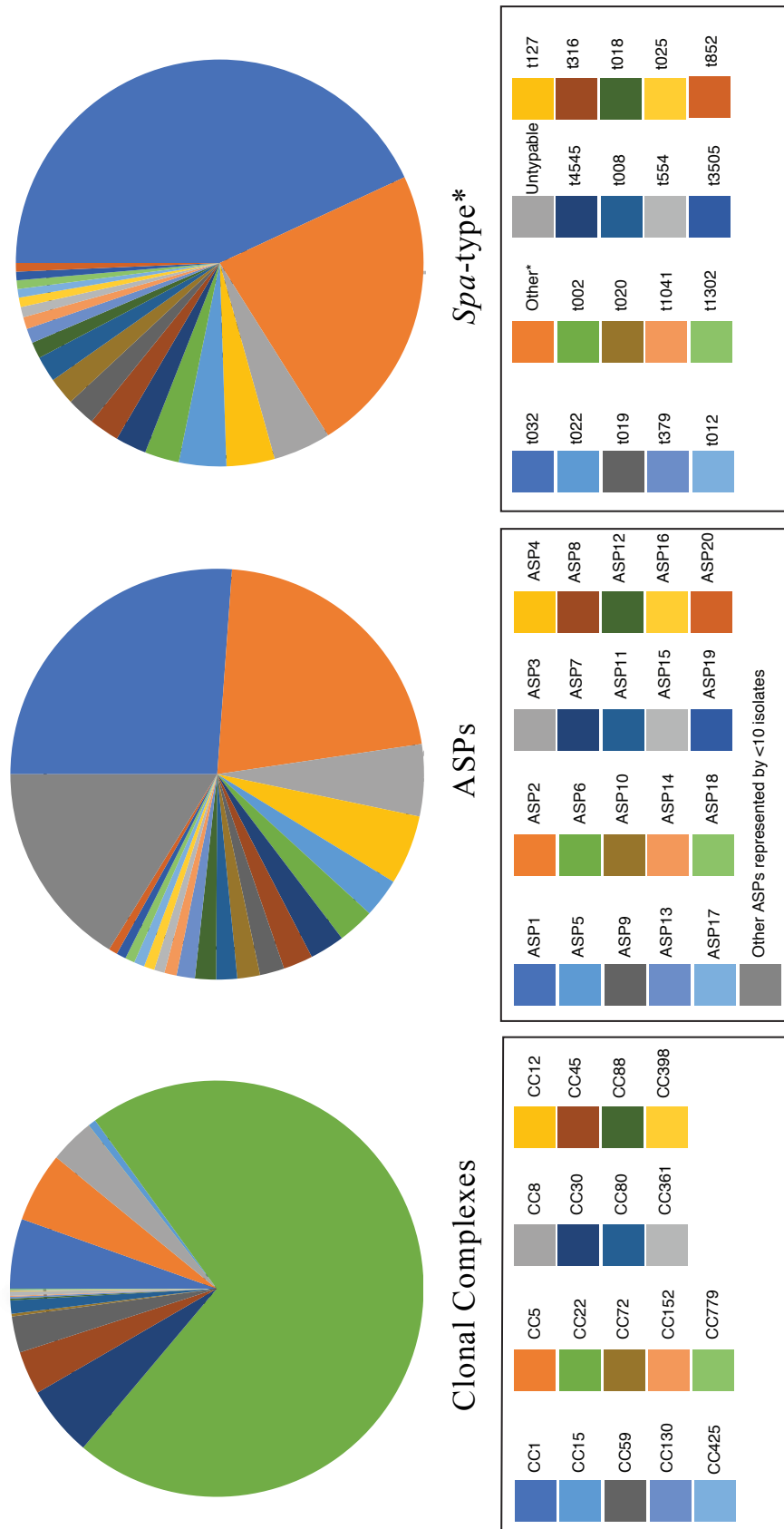
Figure 3.1. Diversity of clonal complexes, ASPs and *spa*-types within the collection.
*For 42 pairs, *spa* comparison was not done.

Table 3.1. Phenotypic ASPs identified within genotypic CCs

CC	Number of Isolates	Percentage of Isolates	Number of Unique ASPs
1	79	5.4	38
5	79	5.4	25
8	51	3.5	23
12	2	0.1	1
15	9	0.6	7
22	1035	71.0	74
30	80	5.5	19
45	49	3.4	11
59	41	2.8	12
72	3	0.2	3
80	15	1.0	6
88	2	0.1	2
130	2	0.1	1
152	1	0.1	1
361	5	0.3	4
398	1	0.1	1
425	1	0.1	1
779	1	0.1	1
1943	1	0.1	1

3.3.1. Accuracy of ASP in Determining Hospital Ward MRSA transmission

Using existing evidence based on the genomic and epidemiological dataset as the reference, this study determined the accuracy of using ASP and a direct ward contact in identifying putative MRSA transmissions as defined above. Comparison of epidemiologically related case-pairs sharing the same ASP was undertaken to define the sensitivity and specificity of this

method (Table 3.2). A total of 2514 case-pairs had direct ward contact. Of these, 403 case-pairs had isolates with an identical ASP. The sensitivity and specificity of ASP in the presence of a direct ward contact was 44% and 85% with a positive predictive value (PPV) of 6.7%. This equates to the investigation of 403 case-pairs in hospital to identify 27 putative transmission events whilst missing a further 35 events.

Table 3.2. Sensitivity, specificity and predictive values of ASP in determining genetic relatedness (<=50SNPs) between different categories of epidemiologically related isolate pairs in the study period.

Combination	No. Comparisons	TP	FP	TN	FN	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Hospital Epidemiological link (Same Ward, Same Time)									
Identical ASP	2514	27	376	2076	35	43.5	84.7	6.7	98.3
Identical ASP, *spa*-match	2346	26	154	2131	35	42.6	93.3	14.4	98.4
Identical ASP excluding ASP1 & ASP2	664	15	46	595	8	65.2	92.8	24.6	98.7
Single ASP mismatch	2514	47	1028	1424	15	75.8	58.1	4.4	99.0
Community Epidemiological link (Same Post code)									
Identical ASP	434	10	102	315	7	58.8	75.5	8.9	97.8
Identical ASP, *spa*-match	390	9	57	317	7	56.3	84.8	13.6	97.8
Identical ASP excluding ASP1 & ASP2	133	6	38	88	1	85.7	69.8	13.6	98.9
Single ASP mismatch	434	16	192	225	1	94.1	54.0	7.7	99.6

3.3.2. Accuracy of ASP in Determining Hospital Ward Transmission when Allowing for a Single Mismatched Antibiotic

Comparison of the similarity of MRSA ASP allowing for a single mismatched antibiotic ('single ASP mismatch') between epidemiologically related case-pairs was undertaken to define sensitivity and specificity (Table 3.2). Comparison to the gold standard demonstrated that 62 case-pairs were confirmed as genetically related, with the sensitivity and specificity of ASP with a single mismatch in the presence of a direct ward contact as 76% and 58% with a PPV of 4%. This equates to the investigation of 1075 case-pairs in hospital to identify 47 transmission events and miss a further 15 events.

3.3.3. Utility of Rare ASPs (Non ASP1, Non ASP2) in Determining Hospital Ward Transmission

Nearly half of MRSA isolates (n=698, 48%) had one of two dominant ASPs (ASP1 and ASP2). Therefore, the data were re-analysed after excluding isolates with ASP1 and ASP2, to determine whether analysis of a match between isolates with rarer, identical ASPs only provided greater accuracy in detecting or excluding putative transmission. After exclusion of the most frequent ASPs, ASP1 and ASP2, the sensitivity and specificity of an identical ASP in the context of a direct ward contact was 65% and 93% (PPV 25%), which equates to investigation of 61 case-pairs to identify 15 putative transmission events whilst missing eight events.

3.3.4. Accuracy of ASP in Determining Transmission in the Community

Coll *et al.* (2017), showed that shared post code acted as a strong epidemiological link for putative MRSA transmission. Therefore, the same methods were applied to those case-pairs with a shared residential post code to determine the accuracy of using ASP and a strong community epidemiological link (shared residential post code) in identifying putative MRSA transmission events in the community. A total of 434 case-pairs had shared residential post code. Of these, 112 case-pairs had isolates with an identical ASP. The sensitivity and specificity of ASP in the presence shared residential post code is 59% and 76% with a PPV of 9%. This equates to the investigation of 112 case-pairs to identify 10 putative transmission events whilst missing a further seven events.

3.3.5. Accuracy of ASP in Determining Community Transmission when Allowing for a Single Mismatched Antibiotic

Comparison to the gold standard demonstrated that 17 case-pairs were confirmed as genetically related, with the sensitivity and specificity of ASP with a single mismatch in the presence of a post code match as 94% and 54% with a PPV of 7.7%. This equates to the investigation of 208 case-pairs to identify 16 transmission events and miss a further one event.

3.3.6. Utility of Rare ASPs (Non-ASP1, non ASP2) in Determining Community Transmission

The sensitivity and specificity of a match between rare, non-ASP1, non-ASP2 in the context of a post code contact was 86% and 70% (PPV 14%), which equates to investigation of 44 case-pairs to identify 6 putative transmission events, whilst missing one event.

3.3.7. *In silico spa*-typing

As laboratory *spa*-typing is a widely accepted and utilised typing method, this study explored the improvement in accuracy of *in silico spa*-typing plus ASP/epidemiology in detecting transmission. The use of *in silico spa-type* in determining genomic relatedness when case-pair isolates had the same ASP and a strong epidemiological link had a sensitivity and specificity is 43% and 93% respectively with a PPV of 14% for hospital cases, and 56% and 85%, respectively, with a PPV of 14%, for community cases.

3.4. Discussion

ASP and epidemiology are widely used in day-to-day infection control practice as indicators of potential MRSA transmission events. Harris *et al.*, 2013, studied an outbreak of MRSA in a special care baby unit. They reported how during the management of this outbreak, MRSA isolates had initially been ruled out of the infection control investigation on the basis of different ASPs but WGS showed them to be closely related. Similarly, Gordon *et al.*, 2017, studied 17 MRSA outbreaks and described how in five of these there were two or more ASPs identified within the groups of genomically related outbreak isolates. Because of the routine use of ASP as a screening tool for MRSA transmission,

yet the uncertainty surrounding the validity of this use, this study aimed to investigate the accuracy of these methods in detecting putative MRSA transmission. The availability of whole-genome data as a gold standard for defining isolate relatedness provides an opportunity to determine the accuracy of the use of ASPs and epidemiology in a way which has not been possible beforehand.

This chapter describes how the sensitivity and specificity of ASP plus an epidemiological link in identifying putative MRSA transmission is low. There is, however, variation in performance when different types of epidemiological links are considered. The first question that the study aimed to answer was, if there are two MRSA positive patients who have been on the same hospital ward at the same time and have the same ASP, what is the likelihood that these represent a putative transmission event? This study showed that the ASP under these conditions had less than 50% sensitivity and a PPV of 6.7%. However, when isolates with the two most dominant ASP were removed from the analysis, the sensitivity and specificity of ASP and epidemiology in identifying transmission within the hospital increased. Indeed, sensitivity and specificity of a strong hospital ward epidemiological link increased to 65% and 93%, respectively, and the PPV increased to 25%. This aligns with past work which suggests that ASP is most useful as a surrogate to identifying outbreak isolates if it is unusual within the particular healthcare setting.[114,116] The performance of ASP and epidemiology when using rare ASPs only was strongest in the hospital setting, identifying the greatest proportion of putative transmissions in the case-pairs investigated.

Secondly, when considering patients who share a residential post code, what is the likelihood that these two MRSA isolates represent a putative transmission? Here, an ASP

match gave a sensitivity of nearly 60%, with a PPV of 9%. However, when excluding isolates with ASP1 & ASP2, the sensitivity, but not the specificity, increases, reaching a PPV of 14%.

It is commonplace in clinical practice to allow for one antibiotic to vary when considering if ASP similarity indicates potential transmission in epidemiologically linked patients. When considering case-pairs linked by a strong hospital epidemiological link, this increased the sensitivity to over 75% but because more case-pairs are potentially linked, specificity is reduced. Importantly, in the hospital setting, over 1000 case-pairs would need to be investigated to identify less than 50 putative transmissions, making this strategy impractical. A similar trend was seen in the community setting, albeit on a smaller scale.

For both the hospital and community setting, when compared to ASP/epidemiology alone, the addition of *in silico spa*-type offered a slight improvement to the PPV of determining putative MRSA transmission events, meaning this may be useful addition to a screening tool, especially in the hospital setting. However, the small increase in performance of the test may not outweigh the additional time and cost that undertaking *in silico spa*-typing of isolates would require.

A number of limitations in this analysis are acknowledged. First, genomic relatedness has been defined based on a SNP cut-off of 50 SNPs or less. Whilst evidence based,[145] the use of any cut-off value may over/under-detect some linkage between isolates. ASPs were determined using the VITEK 2 instrument, technology which may only be used to determine antibiotic susceptibilities for a subset of isolates processed in clinical laboratories, and provide an ASP on 19 antibiotics rather than the six/twelve antibiotics

which may be tested in clinical laboratories at present. However, in the UK, the VITEK2 is increasingly available. Furthermore, using only six antibiotics is likely to decrease discrimination and be even less reliable than any putative ASP links between case-pairs based on 19 antibiotics, therefore the sensitivity, specificity and PPVs above will be greater than that achieved in a setting not using VITEK 2.

Despite these limitations, this study provides much needed insights into the accuracy of epidemiology combined with ASP, routinely used methods in practice, in defining putative MRSA transmission. Whilst these methods are readily available, this systematic analysis demonstrates that compared to the gold standard of WGS plus epidemiology, none of these methods reliably identify/refute transmission events. Improved IT systems in laboratories could increase detection of epidemiological links and ASP matches which may not be noted by practitioners alone, meaning that detection could be optimised. However, given the overall low sensitivity and specificity, in the UK setting, in order to reliably determine a greater proportion of MRSA transmission events, more discriminatory testing such as WGS is necessary. Until universal WGS is commonplace, further work in this area should aim to evaluate the role of targeted WGS to identify those transmissions which are missed through other identification approaches, the false negatives in this study.

This study provides further evidence of the limitations of current practice, and the potential benefits of routine WGS to improve the study of disease transmission and to optimise the control of infection in hospitals.

Chapter 4. Systematic Genomic Surveillance for Potentially High-risk MRSA Lineages in the East of England using WGS

4.1. Surveillance for Potentially High-risk MRSA Lineages

4.1.1. Introduction

In the UK, whilst the vast majority of MRSA infections continue to be caused by the dominant hospital-associated lineage, ST22,[65] some MRSA lineages are considered potentially high-risk. It is difficult to strictly define a high-risk lineage, however those which display resistance to a number of therapeutic agents, a predisposition to cause severe clinical disease (either widespread or in a particular population group) or high transmissibility would be of concern. In the East of England, two of the lineages that may be considered high-risk are USA300 and ST2371. USA300 is considered a high-risk lineage globally because of its rapid spread through the USA and its ability to cause disease in otherwise healthy individuals. The ST2371 MRSA lineage, in contrast, caused a large outbreak locally in both the hospital and the community in this region, in a particularly vulnerable population group (neonates).

A central role of infectious disease surveillance systems is the detection and monitoring of the emergence of novel pathogens or lineages. Chapter 3 (Evaluation of ASPs to detect MRSA transmission) used a genomic gold standard to show how routine practice, using ASP and epidemiology to detect MRSA transmission events has poor sensitivity and specificity. Following on from this, Chapter 4 (Regional surveillance for high-risk MRSA lineages) reports the findings of systematic genomic surveillance for USA300 and ST2371 within a major diagnostic microbiology laboratory serving the East of England to

demonstrate how whole-genome sequencing can be used to investigate the epidemiology of two, potentially high-risk MRSA lineages that otherwise remain largely unmonitored.

4.2. Specific Methods

Both studies used the isolates sourced from the Cambridgeshire MRSA prospective study, sampled and processed as described in Chapter 2 (Materials and Methods).

4.2.1. DNA Sequencing, Phylogenetic Analysis, Definition of Clones

MLST sequence types were identified from the sequence data and the MLST database (http://saureus.mlst.net/), and assigned to clonal complexes as described in Chapter 2 (Materials and Methods). CC8 and ST2371 isolates were identified, and formed the basis for the remaining analyses. Using methods as described in Chapter 2 (Materials and Methods), CC8 isolates were mapped using SMALT to the *S. aureus* USA300 genome FPR3757 (Genbank accession number CP000255.1), and ST2371 isolates to the *S. aureus* E-MRSA reference genome EMRSA-15 (Genbank accession number NZ_CP007659). Mobile genetic elements, insertions and deletions were excluded and SNPs in this core genome were used to create maximum-likelihood phylogenies using RAxML, as described in Chapter 2 (Materials and Methods). Bacterial DNA sequences were deposited in the European Nucleotide Archive (ENA), https://www.ebi.ac.uk/ena, and accession numbers, are provided in Appendix 4.1.

4.2.2. Identification of Isolates

Genome sequence data for 348 USA300 isolates reported previously by Uhlemann *et al.* were sourced from the ENA.[60] The study CC8 isolates were contextualised against the Uhlemann *et al.* collection to characterise USA300 isolates phylogenetically based on their position within the combined phylogenetic tree. Genome sequences from 45 ST2371 isolates from a prior outbreak in Cambridge, 2011,[114] were sourced from the ENA. These were combined with the study ST2371 isolates to determine relatedness.

4.2.3. Accessory Genome

The SCC*mec* subtype was determined using *in-silico* PCR and published primers.[152] For both collections, presence or absence of PVL (*lukF-PV* and *lukS-PV*) was determined. The following methods were applied to the USA300 isolates only. The corresponding sequences for the arginine catabolic mobile element (ACME) and *S. aureus* pathogenicity island 5 (SaPI5) were extracted from the mapping alignment. Mutations in the *cap5* locus were retrieved from the assemblies by comparison to the *cap5* locus of the positive reference strain, Newman, and USA300 reference strains, TCH1516 and FPR3757, as described previously.[153] The presence or absence of acquired genes and SNPs conferring resistance against macrolides, lincosamides, ketolides and quinolones were determined as previously described,[154] with a list of genes given in Table 4.1. For SNPs causing resistance in chromosomal genes the standard mapping and SNP calling approach was used as described in Chapter 2 (Materials & Methods).

Table 4.1. Antibiotic resistance genes

Antibiotic Class	Genes
Quinolones	*grlA*
	grlB
	gyrA
	gyrB
Macrolides and Ketolides	*msrA*
	msrD
	ermA
	ermB
	ermC
	erm33
	ermT
	mefA
	mefE
	ereA
	ereB
Lincosamides	*linA*
	linB
	linC

4.3. Results

4.3.1. USA300 Study

4.3.1.1. Description of Cases Positive for USA300

This study aimed to determine the prevalence of USA300 MRSA in the Cambridgeshire prospective MRSA study (Appendix 2.1). A total of 56 USA300 isolates from 24 cases were identified, giving a case prevalence among MRSA carriers of 1.6% (24/1489). The majority of USA300-positive cases were young (median 32 years, range $3 - 84$ years, interquartile range 25 - 57 years), with a male predisposition (16/24, 67%). The majority of first positive samples were submitted from first-opinion services (ED (8/24, 33%) and GPs (5/24, 21%)). USA300 was identified from carriage screens alone in half of cases (12/24, 50%). Ten cases had skin or soft tissue infection, with or without associated carriage, and two cases had invasive infection. Twenty-two cases resided across the East of England, one residence was unknown and one case normally resided in the US (Figure 4.1). Dates of positive samples were distributed throughout the 12-month study period (Figure 4.2). Review of the microbiological records showed that only three of the 56 USA300 isolates were submitted by the diagnostic laboratory for PVL testing, which were subsequently identified as USA300 by the reference laboratory and therefore detected by the national surveillance mechanisms.
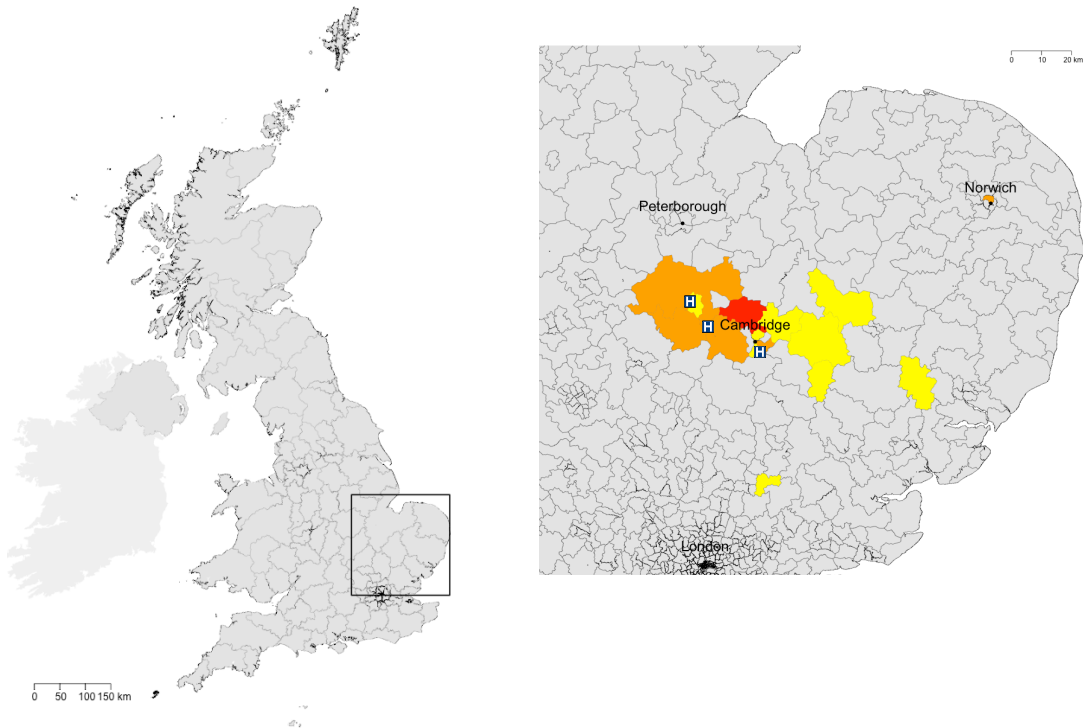
Figure 4.1. Map of East of England, showing the geographical distribution of the residences of patients from which USA300 was isolated, using postcode area. *Yellow: one case; orange: two cases; red: three cases. H: hospital location.*
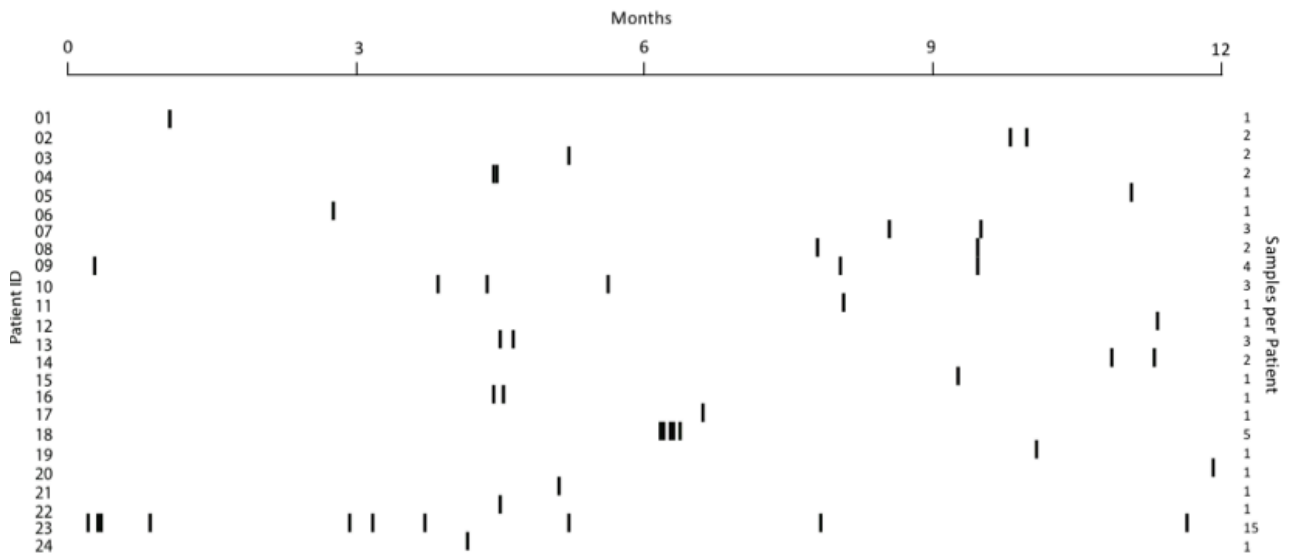
Figure 4.2. Timeline of the date of submission of the each USA300 MRSA isolate per patient between April 2012 and April 2013, showing distribution of isolates over time.

4.3.1.2. Genomic Investigation

A median of one isolate was sequenced per patient (range 1 to 15, a total of 56 isolates). In those cases with more than one MRSA isolate, carriage of non-USA300 was not detected. The median time between the first and last sequenced isolate was 21 days (range 0 to 332 days), and the median pairwise SNP difference between isolates cultured from a given individual was 3 (range 0 to 64). The finding that MRSA USA300 positive cases were distributed in time and space is consistent with numerous independent introduction events. In addition, the phylogenetic tree revealed three groups of very closely related isolates (termed clusters A, B and C, Figure 4.4). Cluster A contained two isolates from two cases (patient (P) P11 and P12) that differed by six SNPs. Samples were submitted three months apart from a GP (ear swab) and a hospital (multi-site MRSA screen) and cases shared the same registered address, suggesting household transmission. Cluster B contained six isolates from two cases (P17 and P18), which differed by a median of 59 SNPs (range 57 to 60), but from other isolates by a minimum of 81 SNPs. The two patients shared a surname but not address, suggesting spread between close contacts rather than household transmission. Cluster C contained 17 isolates from three cases (P22, P23 and P24) with a median difference of six SNPs (range 0 to 9) between isolates from different patients. P22 and P24 had the same registered address and together had a total of 16 MRSA isolates sequenced, the most closely related of which were identical. The single isolate from P23 was also highly related (the closest genetic distance to isolates from P22 and P24 was 1 and 7 SNPs apart, respectively), but a direct or indirect epidemiological link between P22/P24 and P23 could not be identified. The 56 USA300 genomes from this study were combined with 348 MRSA USA300 genomes reported previously and isolated from New York to provide genetic context to the UK isolates. A tree containing all 406 isolates showed that the UK study genomes were interspersed throughout the tree (Figure 4.5).

A:

~ 10 SNPs
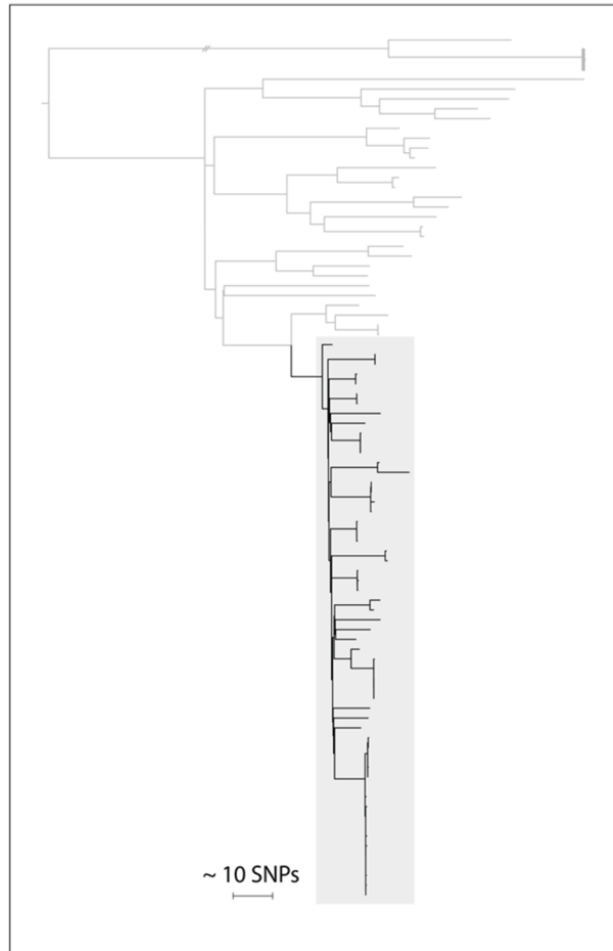
B:
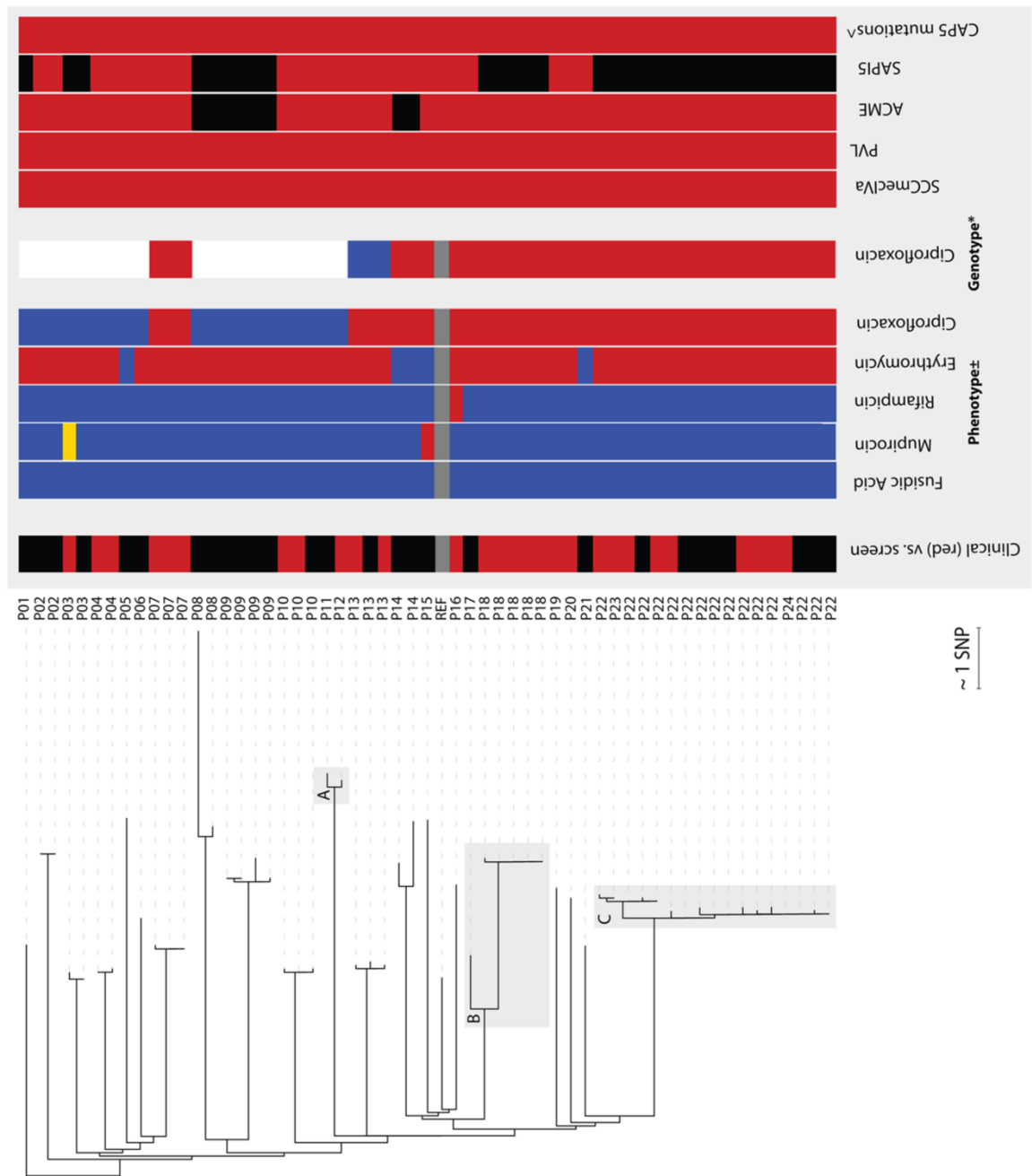
Figure 4.3. A, Midpoint-rooted phylogenetic tree of study CC8 isolates, with USA300 isolates highlighted. 56 isolates residing in the subclade within the grey box were phylogenetically identified as USA300 isolates. B, Detailed USA300 phylogenetic tree rooted on the isolate from P01, with a summary of metadata for each isolate. Person (P) numbers represent the study identifier of each individual from whom the sample was from, with grey boxes indicating pairs or clusters with presumptive epidemiological links.

*±Red, resistant; yellow, intermediate; blue, susceptible. *Ciprofloxacin: red, S80Y and S84L; blue, S80F only; white, none identified; grey, not done. SCCmecIVa, PVL, ACME, SAPI5: red, present; black, absent. ^4 mutations associated with the CAP5 locus identified.*
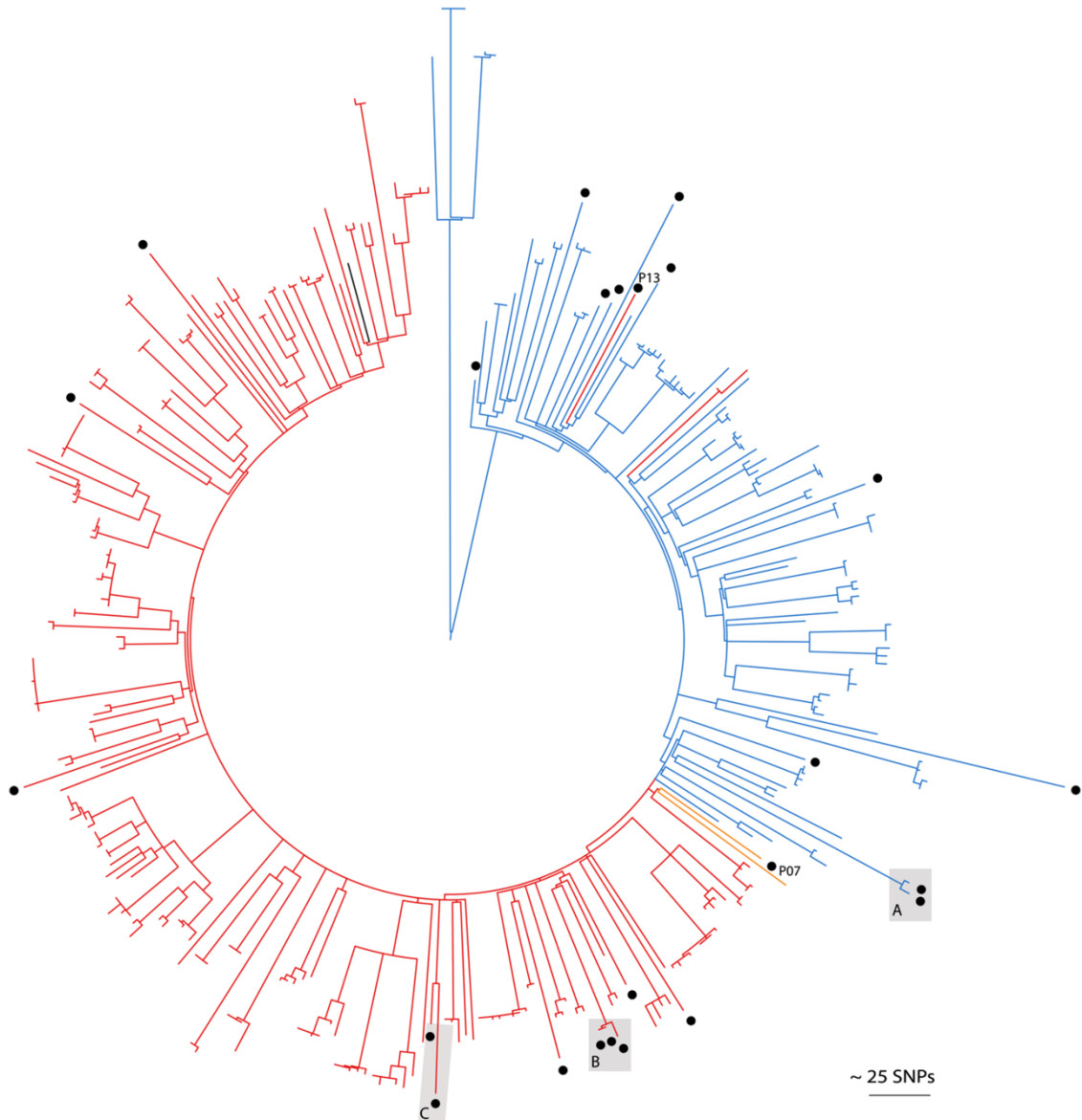
Figure 4.4. Comparison of the first USA300 isolate from each study case (n = 24, circles) relative to previously published USA300 isolates from the United States (n = 348) [6]. Mid-point rooted maximum likelihood tree based on SNPs in the core genome of methicillin-resistant *Staphylococcus aureus* with branch colours representing fluoroquinolone genotypes. Red branches: S80F/Y and S84A/L; yellow: S80F only; blue branches: nil; black branch: reference genome FPR3757. Letters alongside circles indicate epidemiologically linked pairs or clusters.

4.3.1.3. Variability in the USA300 Genome

The presence of mobile genetic elements proposed previously to be associated with USA300 fitness and epidemic spread was investigated (Figure 4.3). Enterotoxins K and Q are thought to enhance pathogenesis through T-cell stimulation and are encoded by genes *sek*2 and *seq*2 within a pathogenicity island, SAPI5.[155,156] SAPI5 was present in 25 of the 56 isolates. The ACME locus is a genomic island associated with USA300 that is composed of at least 33 putative genes and two operons.[157] The *arc* operon encodes genes involved in arginine catabolism, which are important for survival of USA300 in acidic environments.[158] The ACME *speG* gene, which encodes a spermidine acetyltransferase, confers the ability to survive levels of the polyamines spermidine and spermine that are lethal for other strains of *S. aureus*.[159] As described previously,[60] ACME was variably present in the USA300 isolates. Eight isolates from three cases were missing this gene cassette. The dispersed position of ACME-negative isolates in the phylogeny suggested multiple losses of the pre-existing island. ACME-negative isolates did not carry the copper and mercury resistance element (COMER), presence of which would be characteristic of South American strains of USA300-LV.[62] Boyle-Vavra *et al.* recently reported that USA300 failed to produce capsular polysaccharide, which was associated with the presence of four conserved mutations associated with the *cap5* locus when compared with strain Newman.[153] It was confirmed that these four mutations were present in all 56 USA300 isolates. These *cap5* mutations, together with SCC*mec* IVa and PVL, therefore form a consistent marker of USA300 in this study collection, whereas ACME and SAPI5 are variably present.

4.3.1.4. Antibiotic Resistance

The oral antibiotics used to treat MRSA SSTI in the UK and US in single or combination regimens are clindamycin, doxycycline and trimethoprim-sulphamethoxazole, rifampicin, trimethoprim and fusidic acid.[160-162] All 56 isolates were phenotypically susceptible to trimethoprim, and clindamycin (constitutive), and only one isolate tested resistant to tetracycline. None of the 51/56 phenotypically erythromycin resistant isolates tested positive for inducible clindamycin resistance. More than half of isolates (36/56) were phenotypically resistant to ciprofloxacin, and contained known mutations in both *grlA* and *gyrA* (*gyrA* 84L and *grl*A 80Y (n=33), and *grl*A 80F alone (n=3).) Previous studies have reported that USA300 isolates from the US segregated into two clades based on fluoroquinolone resistance genotypes (with or without *gyrA* 84L/*grlA* 80Y/F mutations). [60,61] When the study isolates were considered in the context of the US isolates (Figure 4.4), this was replicated for isolates from 22 cases; the two exceptions being isolates from P07 and P13. Three isolates from P07 carried the 84L/80F mutations and tested phenotypically resistant but resided within the 'susceptible' clade. Three isolates from P13 resided at the top of the resistant clade and tested phenotypically resistant but with an 80F mutation within *grl*A (Figure 4.4).

4.3.2. ST2371 Study

This study aimed to determine the prevalence of the locally problematic MRSA clone, ST2371, in the Cambridgeshire prospective MRSA study (Section 2.1, Materials and Methods). Ten isolates identified as ST2371 were cultured from samples submitted by general practitioners (n=7) and hospital wards (n=3) between June 2012 and February 2013. The ten isolates were cultured from five individuals (Cases A to E), giving a case

prevalence among MRSA carriers of 0.3% (5/1489). Cases were defined as those patients with an isolate identified as ST2371. All ten study isolates were positive for the PVL genes.

4.3.2.1 Phylogenetic Relatedness

The relatedness between ten ST2371 isolates from five individuals sampled in 2012-2013 and 45 isolates from 26 cases of MRSA from an outbreak linked to a local SCBU in 2011 was then investigated. To this end, the ten study isolates were phylogenetically compared to the 45 isolates from the original 2011 outbreak. This comparison demonstrated that the isolates were highly related with a minimum of 2 and maximum of 24 SNPs between isolates from the study and 2011 outbreak (Figure 4.5). The close phylogenetic relatedness of the study isolates to the 2011 outbreak isolates triggered an epidemiological investigation to determine whether epidemiological links could be drawn between the study cases and cases from the previous outbreak.
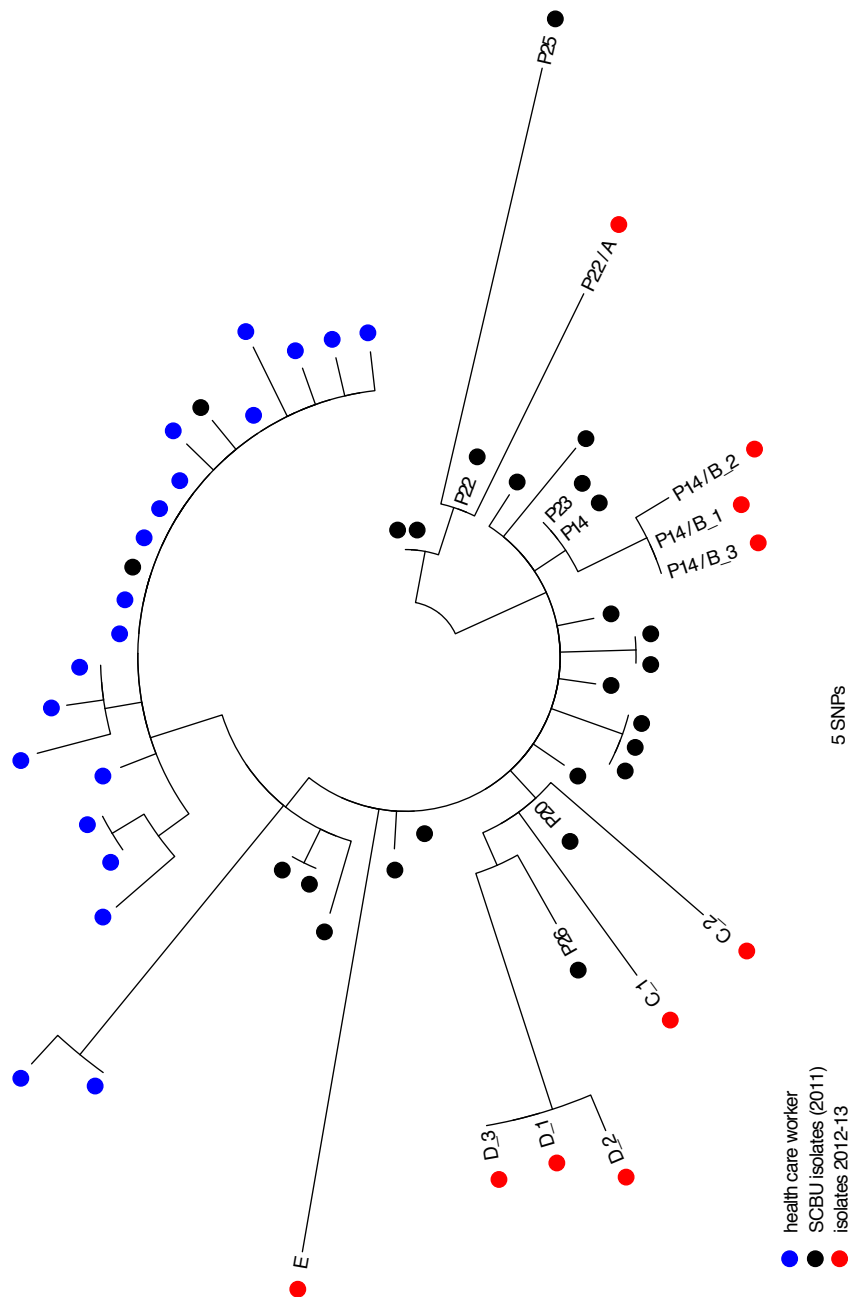
Figure 4.5. Midpoint-rooted phylogenetic tree based on SNPs in the core genome of MRSA. Isolates were mapped against the EMRSA-15 reference genome. Red dots denote isolates from 25 patients and family members investigated during an outbreak on a SCBU in 2011. Blue dots denote 20 individual colonies picked from a nasal swab culture taken from a healthcare worker during the same outbreak. Black dots represent ten isolates from five people who had microbiological samples taken between 2012 and 2013. The numbers shown are the original study number used during the SCBU investigation (prefixed by a P), and the letter A-E for each of the 2012-2013 cases. In the event that more than one sample was taken from the same patient in 2012-2013, these are shown as underscore followed by the sample number (e.g. the second sample from case B is shown as B_2).

4.3.2.2. Descriptive Epidemiology

Investigation showed that all cases had soft tissue infection. Of the five cases, two were female, with age ranging from 1 – 57 years (median: 19 years). None of the cases resided within the same 6-digit post code area and none of the cases were registered to the same GP practice. None of the patients had a history of overlapping hospital admission in the preceding year (Figure 4.7). However, on further investigation, all of the cases had a direct or indirect link to the original outbreak in 2011. The isolation of ST2371 from two of the cases suggest persistence of ST2371 in the community. Cases A and B were part of the original SCBU outbreak (P22 and P14 on the postnatal ward and SCBU, respectively). After being defined as a case in the 2011 outbreak, Case A remained MRSA positive into the early part of 2012, with positive samples in January and March 2012. No further samples were submitted until the sample which was again identified as ST2371 as part of the Cambridgeshire MRSA prospective study, in February 2013. After originally testing positive in the 2011 outbreak, samples from Patient B were not available until October 2012, when ST2371 was again confirmed.

Evidence of familial transmission in the original outbreak is further supported by the spread of this strain between P20, P26, case C and case D. Case C was born at the CUH and was not screened for MRSA before discharge from hospital, but both parents became involved in the SCBU outbreak (P20/P26). Based on a matching surname, case D was a member of the same extended family as P20, P26 and case C.

Case E was born at the CUH and discharged when five days old, which was two days before the birth of the presumed index case of the original SCBU outbreak. The first ST2371 isolate from case E was isolated in 2013 when they were almost two years old, over a year

after the original outbreak was considered over. It is not possible to determine whether acquisition occurred when born at the CUH or from subsequent contact with unidentified carriers in the community.
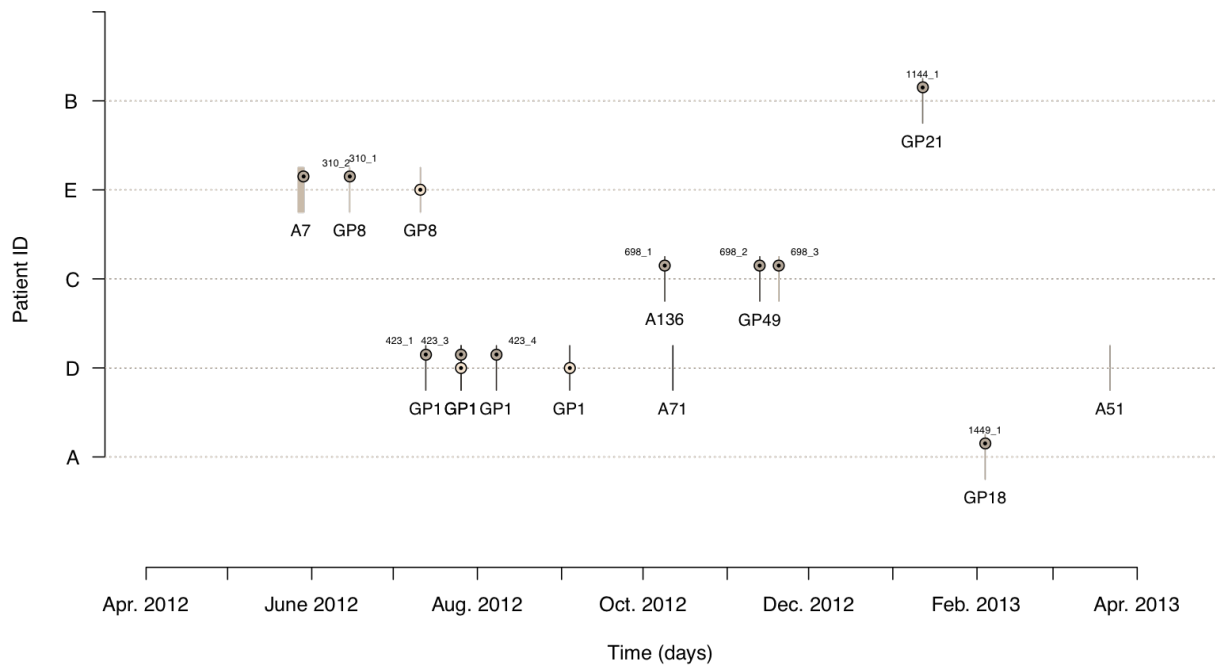
Figure 4.6. Chart displaying timing of visits to hospital A and GP visits for each case during April 2012-2013. Numbers following A (hospital A) or GP (GP surgery) indicated an anonymised code for each ward or GP surgery.

4.4. Discussion

The USA300 study represents the first prospective genomic surveillance study for USA300 in the UK. The current methods for surveillance of invasive staphylococcal infections in the UK are inadequate to monitor potentially high-risk lineages such as USA300. Phylogenetic analysis has indicated a process of repeated introduction throughout the Americas prior to the epidemic.[62] These data suggest a similar pattern, and provide weight for the need to undertake prospective systematic surveillance to define the trajectory of USA300 in the UK. In contrast to when USA300 rapidly disseminated in the USA over 10

years ago, it is now feasible to implement comprehensive, systematic genomic surveillance strategies to guide interventions. Had this been possible during the initial stages of the USA epidemic, an aggressive search and destroy policy or alternative strategy to limit spread may have been implemented.

Systematic prospective surveillance is now required to determine whether this rate changes over time in the UK. Existing studies are sparse and limited by sampling methodology,[64] but it appears that USA300 may be increasing in England. Molecular testing in a national reference laboratory study identified 40 likely USA300 isolates (CC8 SCC*mec*IVa, *spa* t008, *agr* group 1, PVL positive) from across England and Wales over a two-year period (2004-05).[163] The identification of 60% of this total in one rather than two years and from a single region of England indicates a gap in current surveillance knowledge.

Similarly, the ST2371 study demonstrates how, despite causing clinical disease in patients already identified as at risk, a biologically successful lineage can persist within a region and remain undetected by existing surveillance systems. Clinical management of the 2011 ST2371 outbreak included decolonisation recommendations, but this failed to control the persistence of this clone of MRSA in the community. Decolonisation was likely ineffective in cases A and B. Explanations for this include not implementing or completing the course of decolonisation, failed decolonisation, or limiting decolonisation to individual family members followed by repeated household acquisition. Two individuals involved in both the original outbreak and surveillance study experienced disease for at least 15 months after their initial outbreak samples.

Despite causing a clinical outbreak within a vulnerable population in the SCBU at CUH, this lineage has not been monitored since. By demonstrating that the lineage continued to cause disease, and even persisted in patients who were linked to the original outbreak, this shows that there is both a clinical and public health need to review how such lineages are monitored and managed in the longer term. Both studies provide evidence of how using systematic WGS can provide actionable epidemic intelligence of potentially high-risk clones.

Chapter 5. Whole-Genome Sequencing to Investigate Cryptic Community Transmission of ST22 MRSA

5.1. Introduction

At the time of writing, only 40% of bacteraemia cases reported to PHE are attributed to a hospital, which suggests that transmission outside of hospitals is a substantial contributor to overall MRSA bacteraemia rates.[164] Definitive evidence for community transmission as a driver of MRSA infection in the UK is limited, but is supported by a recent epidemiological and bacterial genomic survey that captured transmission events over a 16,000km^2 area of the East of England.[145] This argued against the conventional wisdom that ST22 is largely healthcare associated in the UK,[65,165] and provided evidence for a substantial burden of MRSA transmission outside of hospital settings (i.e. in the community).

In addition to identifying high-risk lineages, as shown in Chapter 4 (Regional surveillance for high-risk MRSA lineages), infection surveillance also aims to detect lineages which occur in larger than expected numbers or within unusual or unexpected locations. In the Cambridgeshire prospective MRSA study (Appendix 2.1), isolates less than 50 SNPs apart were grouped into clusters. This approach granted the opportunity to investigate cases in the 'opposite direction' to a traditional infection control investigation, identifying linked cases via the relatedness of the bacteria prior to exploring epidemiological links. In routine MRSA investigations it is more typical to identify those patients who are epidemiologically linked and then to source samples from each case for typing to determine bacterial relatedness.

This chapter builds on this and uses the high-resolution provided by WGS to characterize the community-based transmission of the typically nosocomial lineage, EMRSA-15, in a GP surgery.

5.2. Specific Materials & Methods

5.2.1. Study Design

A cluster of 13 MRSA-positive individuals registered to a single GP surgery in Cambridgeshire was first detected during the Cambridgeshire prospective MRSA study, as described in Chapter 2 (Materials and Methods) In brief, 1,465 individuals were identified with MRSA isolated at least once from either screening swabs and/or clinical specimens, and WGS of 2,282 MRSA isolates from these cases. Combined analysis of WGS data revealed a single large cluster of closely-related MRSA (defined based on a pairwise SNP distance <50) that contained 22 isolates from these 13 individuals. This formed the starting point for a public health investigation and the study described here.

5.2.2. Public Health Investigation

The detection of the MRSA cluster resulted in an investigation conducted in May 2015 with the local PHE health protection team. The GP surgery had more than 10,000 registered patients and provided specialist services including diabetic and podiatry clinics. The 13 people involved in the MRSA cluster (defined as cases) were sent an information sheet and details of opt-out consent prior to individual GP record review (Appendix 5.1). If consent was not withheld and records were available, data were collected on demographics, co-

morbidities and date of first MRSA detection using a data collection proforma (Appendix 5.2). In the six months prior to each patient's first recorded positive MRSA result, healthcare attendance (primary care, hospital outpatient, or in-patient) and microbiological samples that were MRSA-negative were recorded.

Incidence rates of MRSA-positive individuals were calculated per 10,000 registered patients at the study surgery. The CUH laboratory information system was used to determine incidence of MRSA positivity based on samples submitted to CUH from four comparable practices within the same region (defined as practices with more than 10,000 registered patients in the same GP Classification Group).[166] All data were collected and analysed within the context of the public health investigation.

Staff at the GP surgery were invited to undergo MRSA screening (triple site screening consisting of nose, throat and groin swabs) following attendance at an information session and after providing written consent (Appendix 5.3). Environmental MRSA screening was performed at 40 sampling points in the building (Appendix 5.4). Samples were taken from high-contact equipment and surfaces in the following areas: two randomly selected medical clinic consultation rooms, two nursing clinic rooms (where the ulcer clinic, which was the strongest epidemiological link between patients, was held), and shared patient waiting areas. At each sampling point, an area of approximately 10cm-by-10cm (or entire surface of handles) was swabbed and cultured for MRSA using direct plating onto chromogenic agar.[167]

Extended case-finding was performed to identify further cases that might be linked to the cluster over a longer time period, and for whom MRSA isolates had been stored and could

be retrieved for sequencing. This involved three different approaches. (i) A retrospective search was performed of the CUH information system for MRSA-positive samples submitted by the GP surgery between January 2006 until June 2015. These data were then cross-referenced with the bacterial archive database to determine if isolates had been stored at -80°C. (ii) Laboratory surveillance was conducted in the laboratory between November 2015 and February 2016 to detect MRSA-positive individuals from the GP surgery. (iii) Recent PIRs at the GP surgery were reviewed. Isolates were requested from the receiving hospital for WGS and patient records reviewed as described above.

5.2.3. WGS, Typing and Data Analysis

DNA was extracted, libraries prepared, and 150-bp paired end sequences determined on an Illumina HiSeq2000 (original study isolates) or MiSeq (isolates identified through additional case finding). Methods were as described in Chapter 2 (Materials and Methods). Details of reads, depth of coverage/N50 are provided in Appendix 5.4. Sequence data were submitted to the European Nucleotide Archive (www.ebi.ac.uk/ena), accession numbers are also listed in Appendix 5.5. STs and CCs were assigned using sequence data as described in Chapter 2 (Materials and Methods). Isolates were mapped using SMALT to the E-MRSA15 reference genome (strain HO 5096 0412, accession HE681097) and the phylogenetically informative core genome was identified for each isolate. SNPs were used to create a mid-point rooted, maximum-likelihood phylogeny using RAxML with 100 bootstraps, and trees were visualised as described in Chapter 2 (Materials and Methods). *In-silico* PCR of the variable X-region of the *spa* gene was undertaken using the genome data and published primers.[151] *spa*-type was then determined using spaTyper (http://spatyper.fortinbras.us).

5.3. Results

During the year-long prospective MRSA study of carriage and clinical MRSA isolates in the Cambridgeshire between April 2012 and April 2013, a number of potential outbreaks based on genomic relatedness and epidemiological links were identified (Appendix 2.1). The largest potential outbreak consisted of thirteen MRSA-positive individuals (22 isolates) registered with the same GP surgery in Cambridgeshire and therefore was of particular interest. All thirteen isolates were ST22 and part of the EMRSA-15 clade (Figure 5.1). An investigation was initiated to rule out on-going transmission, and to elucidate if this represented community-based transmission or 'spill-over' from a more likely source such as hospital or LTCF.

Extended case finding identified additional MRSA-positive individuals attending the same GP with samples available for sequencing (Figure 5.2). First, retrospective review of electronic laboratory records identified four individuals with a total of seven isolates retrievable for sequencing, one of whom (Patient(P)04) had already been identified in the initial 13 cases. Second, prospective surveillance of MRSA-positive samples sent from the GP surgery over three months between November 2015-February 2016 and surveillance of new positive MRSA samples by the infection control team identified three retrievable isolates from three individuals. Third, two PIRs had been undertaken in 2014/2015 (P12/P13). Both patients had died with MRSA bacteraemia in another regional hospital. A single isolate from each blood culture was obtained for each patient from the admitting hospital. A summary of the 22 patients (34 isolates) from the original study and additional case-finding is provided in Table 5.1. The median number of MRSA isolates per patient was 1 (range 1-4). Four patients had only screening samples submitted. Of those clinical

samples submitted, 61% were reported as superficial swabs of lower limbs/foot, while three were from blood cultures and one from pus (all from different patients.)

~ 100 SNPs
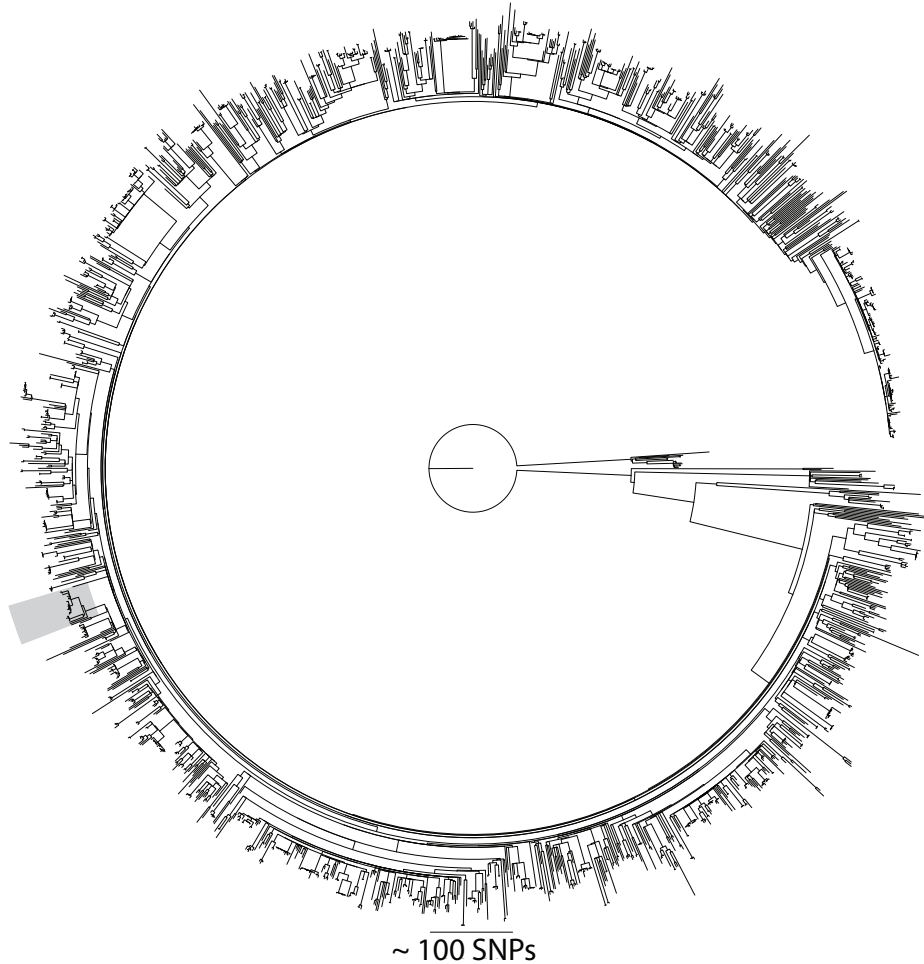
Figure 5.1. Maximum likelihood tree generated from SNP sites in the core genome for 1,715 CC22 isolates from the 2012-2013 study. The clade highlighted in grey is the largest cluster (with a maximum SNP cut-off of 50) within the collection, and represents patients registered to the study GP surgery.
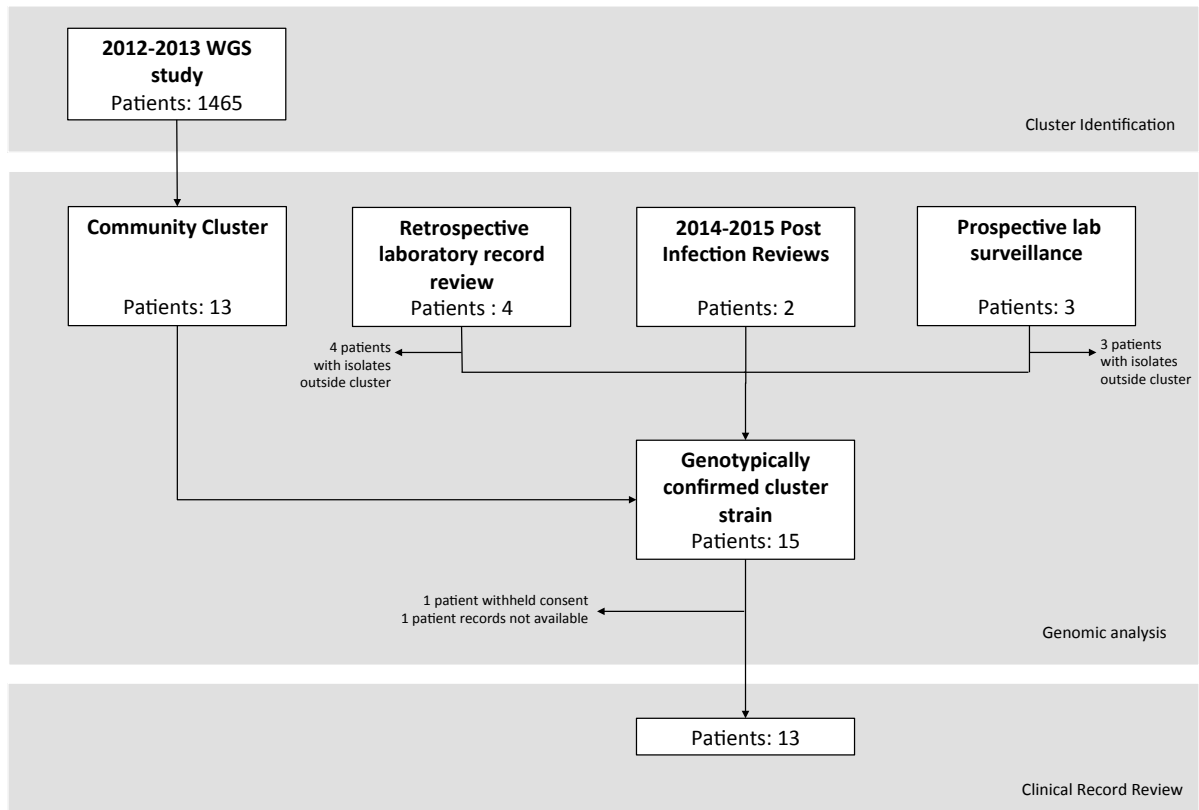
Figure 5.2. Flow diagram summarising the identification of study patients. One patient was captured by both the community cluster and extended retrospective laboratory record review.

Table 5.1. Patient and sample information.

Study ID	Isolation year	Sample type/site	Method of identification	MLST type	Within phylogenetic cluster?	Included in public health investigation?
P01_1	2012	Clinical, foot	Coll et al., 2016	22	Yes	Yes
P02_1	2012	Clinical, leg	Coll et al., 2016	22	Yes	Yes
P02_2	2013	Clinical, leg	Coll et al., 2016	22	Yes	Yes
P03_1	2012	Clinical, foot	Coll et al., 2016	22	Yes	Yes
P03_2	2012	Screen	Coll et al., 2016	22	Yes	Yes
P04_1	2012	Clinical, leg	Coll et al., 2016	22	Yes	Yes
P04_2	2015	Clinical, ankle	This study, lab record review	22	Yes	Yes
P04_3	2014	Screen	This study, lab record review	22	Yes	Yes
P04_4	2014	Screen	This study, lab record review	22	Yes	Yes
P05_1	2012	Clinical, unspecified	Coll et al., 2016	22	Yes	Yes
P06_1	2012	Clinical, leg	Coll et al., 2016	22	Yes	Yes
P06_2	2012	Screen	Coll et al., 2016	22	Yes	Yes
P07_1	2012	Clinical, leg	Coll et al., 2016	22	Yes	Yes
P07_2	2012	Clinical, leg	Coll et al., 2016	22	Yes	Yes
P07_3	2012	Clinical, leg	Coll et al., 2016	22	Yes	Yes
P08_1	2012	Clinical, foot	Coll et al., 2016	22	Yes	Yes
P09_1	2012	Clinical, leg	Coll et al., 2016	22	Yes	Yes
P10_1	2012	Screen	Coll et al., 2016	22	Yes	Yes
P11_1	2013	Clinical, leg	Coll et al., 2016	22	Yes	Yes
P11_2	2013	Screen	Coll et al., 2016	22	Yes	Yes
P12_1	2014	Clinical, blood	This study, PIR	22	Yes	Yes
P13_1	2015	Clinical, blood	This study, PIR	22	Yes	Yes
P14_1	2012	Clinical, leg	Coll et al., 2016	22	Yes	No
P14_2	2012	Clinical, unspecified	Coll et al., 2016	22	Yes	No
P15_1	2012	Clinical, back	Coll et al., 2016	22	Yes	No
P15_2	2012	Screen	Coll et al., 2016	22	Yes	No
P15_3	2013	Screen	Coll et al., 2016	22	Yes	No
P16_1	2014	Clinical, genital	This study, lab record review	6	No	No
P17_1	2014	Clinical, finger	This study, lab record review	45	No	No
P18_1	2014	Screen	This study, lab record review	22	No	No
P19_1	2015	Screen	This study, lab record review	45	No	No
P20_1	2015	Screen	This study, prospective surveillance	45	No	No
P21_1	2016	Clinical, blood	This study, prospective surveillance	1539	No	No
P22_1	2016	Clinical, abscess	This study, prospective surveillance	22	No	No

5.3.1. Genomic Analysis

STs were derived from WGS data for the 12 MRSA isolates identified through additional case finding. The predominant ST was ST22 (7 isolates, 3 individuals), the remainder being ST45 (3 isolates, 3 individuals), ST6 (1 isolate) and ST1539 (a single-locus variant of ST221, 1 isolate). The non-ST22 cases were excluded from further analysis. After combining the 22 ST22 isolates from the original study and 7 from additional case-finding to form a total of 29 ST22 isolates, a maximum-likelihood tree was constructed based on SNPs in the core genome compared to the EMRSA-15 reference genome. This demonstrated clustering of 27 of the 29 ST22 isolates from 15 individuals (Figure 5.4), now referred to as cases. Of these, 13 had been identified in the original cluster and the additional two isolates were from two cases (P12 and P13) identified during PIRs of fatal bloodstream infections. The median pairwise SNP distance between the 27 cluster isolates from these 15 cases was 21 (range 0 to 58, interquartile range 10 to 37). The median pairwise SNP distance for cluster isolates from the same person (in the 8 cases with more than one isolate) was five (range 0 to 60, interquartile range 1.5 to 15.5). One patient (P04) had cluster isolates that extended over a period of 34 months (a basal isolate in 2012, and three isolates in 2014/15 with pairwise distances of 60, 59 and 57 SNPs from the 2012 isolate). *spa*-typing showed that the cluster was formed of two main *spa* types (t032, t294) with three additional variants (Figure 5.3).
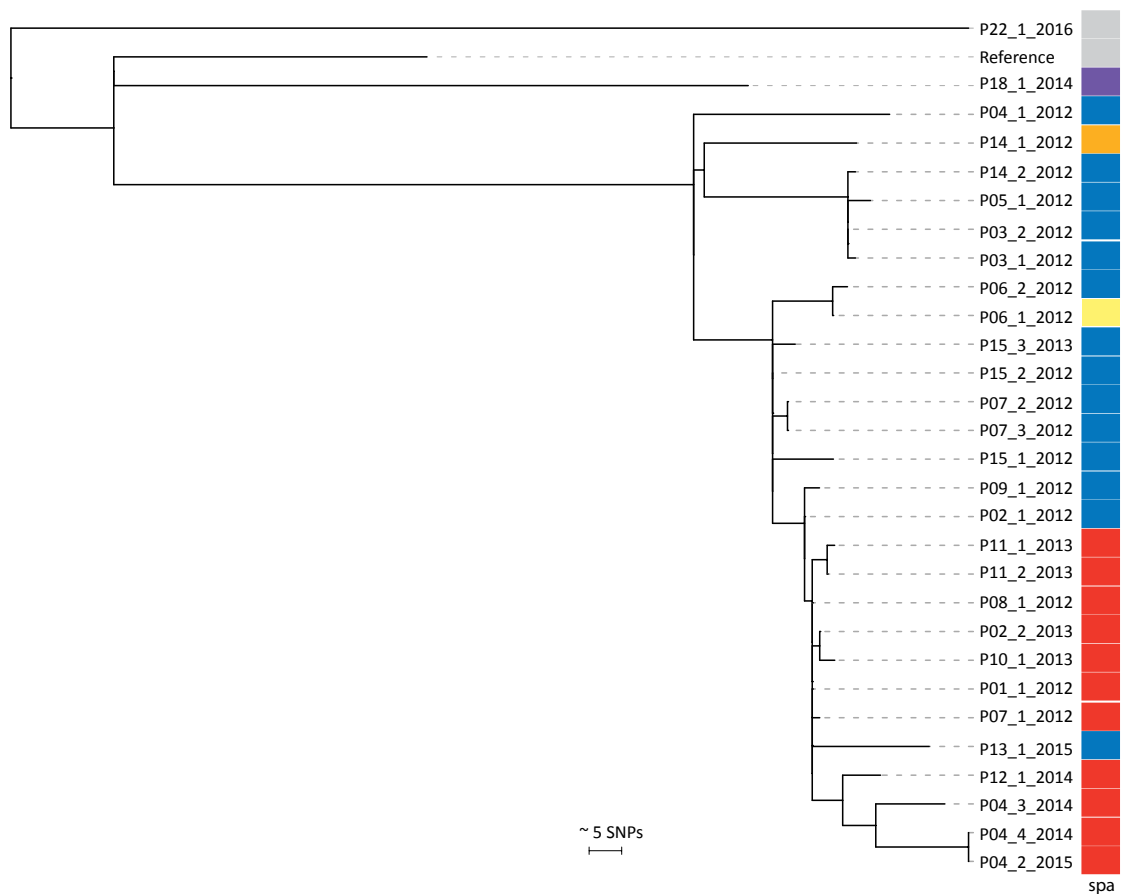
Figure 5.3. *spa*-**type of 29 MRSA ST22 isolates from 15 cases linked to a GP surgery.** Each isolate is labelled as patient (P) study number_isolate number_year of isolation. Midpoint rooted maximum likelihood tree based on SNPs in the core genome. Coloured bars indicate *spa* type: red, t294; blue, t032; yellow, t379; orange, t1302; purple t492; grey, not typable/not done.
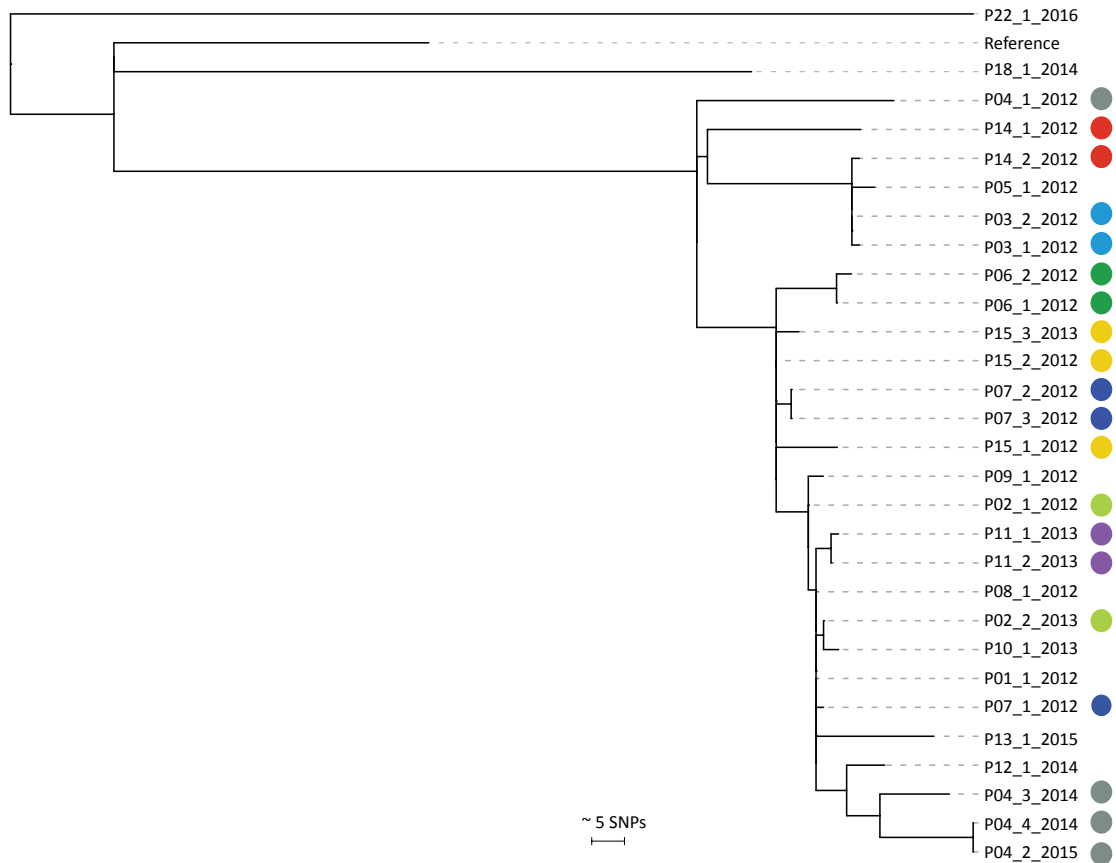
Figure 5.4. Phylogenetic analyses of 29 MRSA ST22 isolates from 15 cases linked to a GP surgery. Midpoint rooted maximum likelihood tree based on SNPs in the core genome. Each isolate is labelled as patient (P) study number_isolate number_year of isolation. Circles indicate multiple isolates from the same patient, with each color being unique to a patient. Cases without circles signify those with a single isolate.

5.3.2. Public Health Investigation

To further understand the cluster, a public health investigation was performed to investigate the 15 genomically linked cases, together with screening of staff and the environment for the presence of MRSA. Two of the 15 cases (P14/P15) were excluded from further epidemiological analysis due to missing patient records or refusal of consent. Of the remaining 13 cases, the median age was 80 years (range 12-91, IQ range: 61.5-81.5) at the time of the investigation, and six cases (46%) were women. Geographical mapping of first MRSA isolation date and place of residence for each case demonstrated that cases lived

within 5.6km of each other and two individuals (P10/P11) lived on the same street. No cases lived in the same household or LTCF.

Review of sample requesting information showed a predominance of lower limb swabs (cases with samples including lower limb, 9; screen alone, 1; bacteraemia alone, 2). GP medical records revealed that the date of first recorded MRSA positive sample for cases ranged from 2006 to 2015 (Figure 5.5A). Healthcare contact by each case in the six months prior to first MRSA detection was extensive for all but two patients (Figure 5.5B). Six of the 13 cases had attended hospital in this period, of whom three cases (P08/P11/P12) had attended only one hospital, two cases (P05/P13) had attended two different hospitals, and one (P10) had attended three different hospitals. Crucially, no overall link could be made between cases and attendance at a hospital (Figure 5.5A). Six individuals had one or more samples that were negative for MRSA in the six-months prior to their MRSA first detection date, and had no record of hospital attendance in the intervening time. Eleven of the 13 cases had attended the GP leg ulcer clinic. P5 and P11 had not, but P11 lived in the same road as P10.
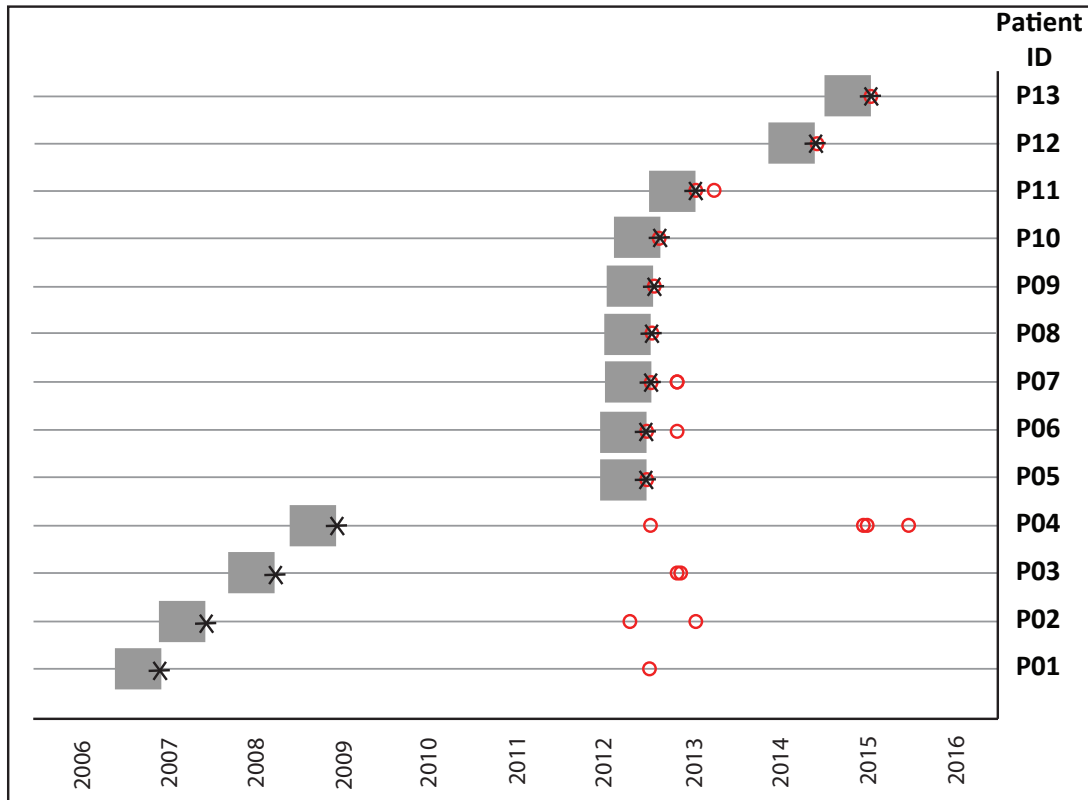
Figure 5.5A. Timeline showing dates of first positive MRSA sample for cases. Date of first known positive MRSA sample (denoted by black star) for the 13 cases investigated in public health investigation, and the preceding 6-month window (grey boxes) during which contacts with healthcare for each case were established. Red open circles denote date of genomically confirmed cluster lineage MRSA samples for each individual.

In 180 days prior to MRSA first detection, number of:

Patient ID	MRSA negative samples	Hospital visits	Ulcer clinic visits	Other primary care visits
P13	-	4	13	14
P12	2	2	37	11
P11	1	1	10	5
P10	-	13	0	13
P09	-	0	31	2
P08	3	6	0	26
P07	1	0	47	1
P06	2	0	27	0
P05	-	2	0	0
P04	-	0	26	2
P03	1	0	23	3
P02	-	0	23	0
P01	-	0	3	3
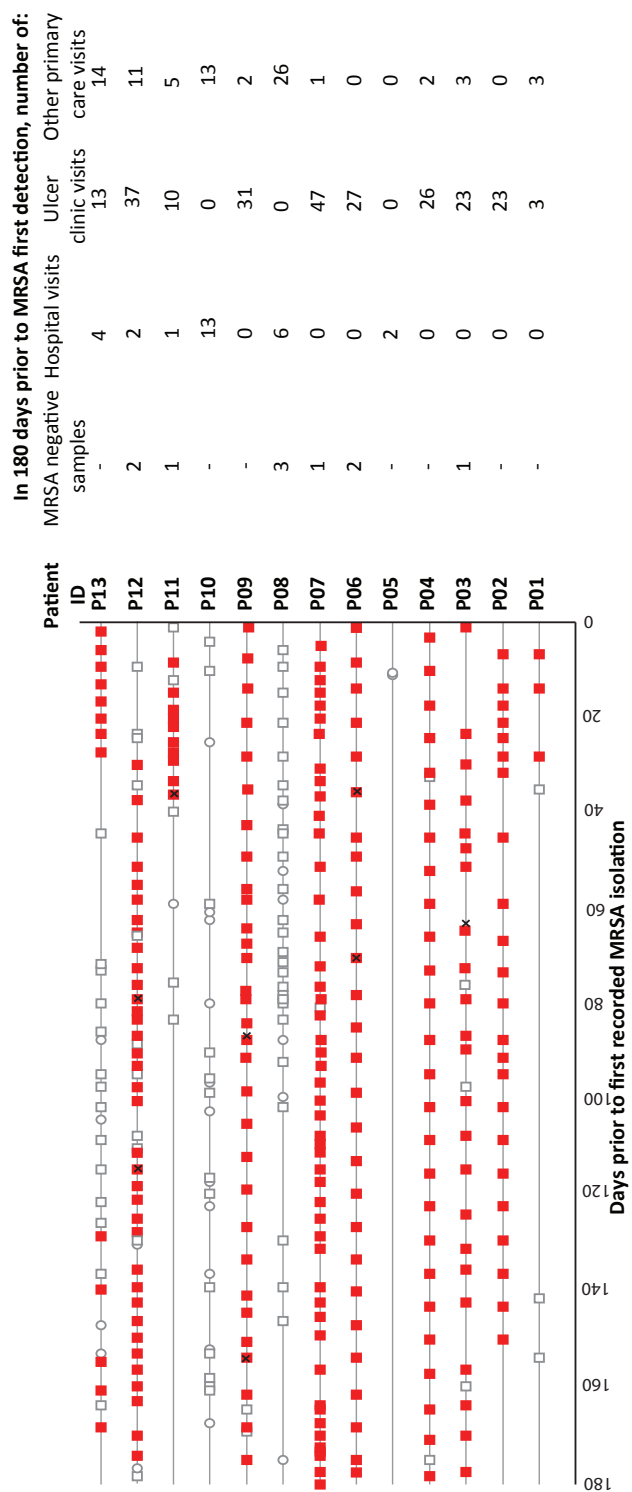
Figure 5.5B. Timeline summarising healthcare contact for 13 cases in the six months prior to first MRSA positive sample. The timeline for each case does not necessarily overlap and ranges between 2006-2015 as shown in Figure 5.5A. Recorded contact with healthcare represented by: Open circle, hospital; Red square, ulcer clinic; Open square, any other GP visit. Black crosses indicate date of negative MRSA sample.

A total of 57 GP surgery staff (approximately 90% of current clinical/non-clinical employees) received multi-site MRSA screens, all of which were negative. This included four nurses who had worked at the ulcer clinic since the first positive MRSA samples in 2008. Forty environmental samples were taken from communal waiting areas and clinic rooms, all of which were also MRSA-negative.

Given that this cluster was only identified by genome sequencing, the question of whether the incidence rate of MRSA-positive samples in the practice had been higher than that expected was explored. This was achieved by comparing the incidence rate of MRSA-positive samples submitted to the CUH diagnostic microbiology laboratory between 2006-2013 between the study GP surgery and four other practices of a similar size and patient demographic within Cambridgeshire. This showed a fluctuating rate over time for all four practices, with no identifiable outbreak signal for the general practice under investigation (Figure 5.6).
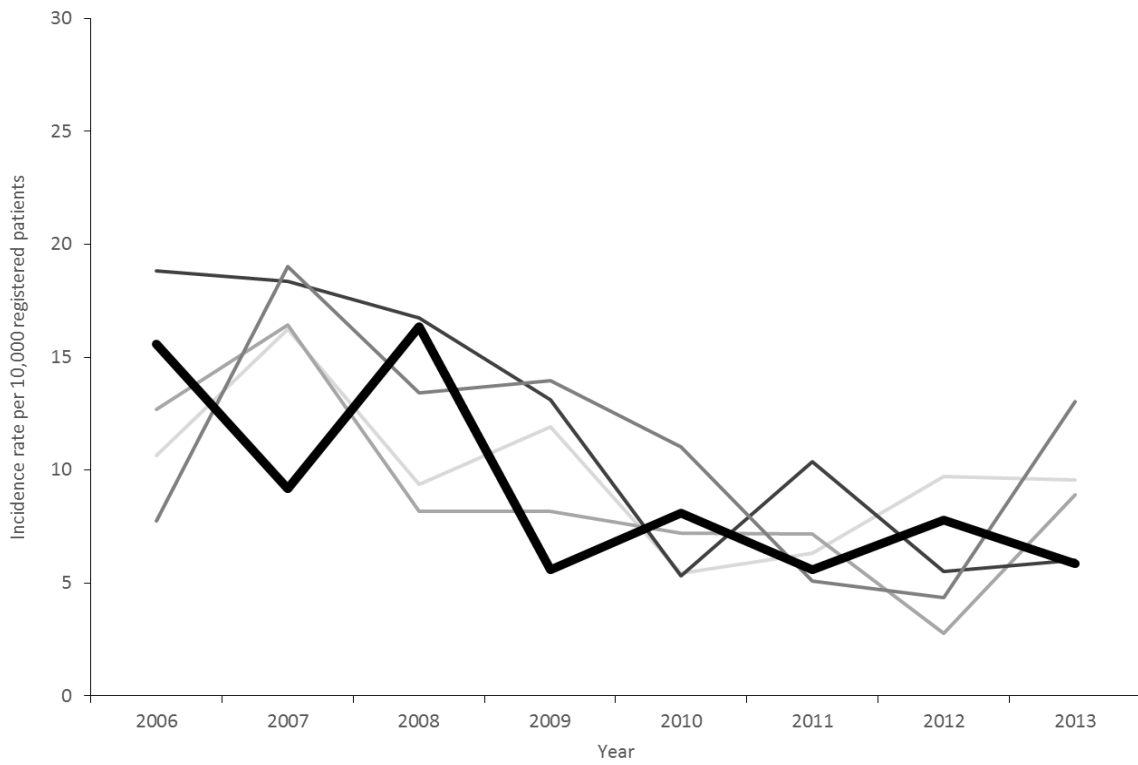
Figure 5.6. Comparative incidence rate of MRSA for the study GP surgery and four comparable general practices. Bold line represents practice studied. Practices not labelled to maintain organisational anonymity.

5.4. Discussion

This chapter describes how routine infection control failed to detect or prevent a community cluster involving fifteen people who carried or were infected with ST22 MRSA. This was despite two fatal cases of bacteraemia that were investigated using standard public health procedures,[107] but were not linked to each other or the cluster until WGS was undertaken. Overall, epidemiological evidence was consistent with onwards transmission of MRSA in the community, although the precise circumstances under which this occurred could not be defined. Most patients were high users of primary care services including a GP leg ulcer clinic, although transmission through other unidentified contacts cannot be ruled out.

One case (P04) had a history of testing MRSA positive since December 2008, and WGS on available isolates confirmed carriage of the same MRSA lineage over a period of 34 months. The diversity within the isolates from P04 encompasses that of isolates from all other cases, potentially suggesting that persistent carriage in this case had contributed to spread of this lineage. Due to limited sampling, it was not possible to rule out re-infection, but the most recent common ancestor of these isolates would have dated to around 2006 (based on a SNP rate of ~3.5 SNPs/genome/year in ST22),[126] consistent with carriage since that date. The important contribution of long-term MRSA carriers to transmission in hospitals has been shown previously,[116,141], and this study shows that it is likely to be relevant in other settings. Decolonisation of persistent carriers with chronic wounds such as leg ulcers is notoriously difficult, and rigorous infection control is required during treatment such as dressing changes when bacterial shedding can occur. MRSA was not isolated from staff or the environment at the GP practice during a point-prevalence survey, but this was performed a considerable period of time after the cluster had become established and was undertaken largely to identify modifiable factors.

MRSA ST22 is the most common MRSA lineage associated with healthcare-associated infection in the UK,[65] and based on the higher overall prevalence of MRSA in hospitals versus the community in the last few decades it has been assumed that the predominant directionality of spread is from hospitals into the community. Previous studies conducted in the UK have isolated ST22 from the community,[168,169] but bacterial typing lacked sufficient resolution to infer transmission. To support this, *in silico spa*-typing of the cluster isolates in this study was undertaken, and based on the presence of a number of *in silico spa*-types it is unlikely that this cluster would have been identified. Chapter 4 (Regional surveillance for high-risk MRSA lineages) describes how WGS was used to confirm that transmission of a PVL-positive, single locus variant of ST22 occurred from a special care baby unit into the community where it subsequently persisted.[114,170] By contrast, the findings of this study suggest that most cases (9 of 13) associated with the MRSA cluster had either not attended a hospital or had at least one intervening sample that was MRSA-negative in the six months prior to first MRSA detection. The majority of cases had links to clinic attendance in the community (in particular for ulcer care), providing genomic evidence for transmission of this typically-nosocomial lineage within community rather than hospital settings.

A greater focus is needed to detect MRSA transmission in the community if overall MRSA bacteraemia rates are to be further reduced. The role of infection prevention and control in the community will become increasingly relevant as National Health Service initiatives are rolled out that increase delivery of care outside of hospitals,[81] and will require a review of the current predominantly hospital-centric structure of infection services.[168] In order to determine if passive surveillance with a defined 'trigger' threshold would have detected

this cluster, rates of MRSA over time for the study practice and other similar practices were compared. The protracted period over which transmission occurred in the cluster described here meant rates of MRSA over time for the GP surgery were comparable with other similar practices. Consequently, the real-time analysis of routine data from this practice is unlikely to have triggered an outbreak investigation. However, the addition of WGS allowed robust assessment of the relatedness of MRSA isolates and the cases involved in the outbreak.

This study has a number of limitations. It cannot be excluded that the outbreak may have been detected through other typing methods not undertaken here, such as pulsed-field gel electrophoresis. Sampling was not undertaken for asymptomatic MRSA carriers in the wider community, which is likely to have under-represented the extent of the cluster. Only a small proportion of the MRSA isolated from samples submitted by the GP surgery were available for sequencing, reducing the number of cases that could be included from the retrospective look-back. The study was not sufficiently powered to conduct a case-case design (cases with MRSA assigned to the cluster versus unrelated MRSA cases) to determine specific risk factors for MRSA acquisition, as comparison between practices was limited due to the variation in services provided. Finally, not all staff who may have been involved in the care of patients within the cluster were screened for MRSA due to staff turnover.

In conclusion, the detection of transmission and outbreaks associated with MRSA ST22 carriage and infection in the community is incomplete. In particular, this study demonstrates the need to consider GP surgeries as well as hospitals and LTCFs as potential transmission hotspots. Whilst WGS of all MRSA isolates from GP surgeries may not be cost-effective, this work demonstrates how universal WGS of bacteraemia isolates can

detect relatedness and potential transmission events in settings which are not typically regarded as foci of transmission. Systematic WGS could provide more accurate attribution of source, provide a mechanism for more efficient targeting of infection control, and lead to further reductions in the number of people who become colonised by, and go on to develop MRSA bacteraemia.

Chapter 6. Genomic Surveillance of MRSA Associated with Bacteraemia in England

6.1. Introduction

Since the national MRSA bacteraemia surveillance programme conducted by PHE was introduced in 2001, it has not included routine submission of isolates for characterisation. Isolates submitted to the PHE Staphylococcal Reference Laboratory are highly selected and are submitted in order to type isolates as part of suspected outbreak investigations in healthcare and community settings, and/or to detect specific genes in isolates from patients with suspected toxin-mediated disease. Strain characterisation is undertaken through *spa* typing, MLST, SCC*mec*-subtyping, toxin gene profiling, and as required, WGS. It is possible that a large amount of information regarding the population structure of disease-causing MRSA in England may have been missed as a result of using this *ad hoc* isolate collection for surveillance.

National bacteraemia surveillance was originally introduced in England to compare MRSA rates between hospitals and then enhanced to aid direction of clinical interventions.[38] Combined with comprehensive sampling regimens, WGS technologies now provide the opportunity to study the natural history of successful MRSA clones at great resolution, and identify clonal expansions to monitor in case of widespread dissemination.[171] This chapter describes a proof-of-principle study to determine the feasibility and potential benefits of combining prospective epidemiological and genomic surveillance of MRSA bacteraemia on a national scale within a public health organisation.

6.2. Specific Methods

6.2.1. Study Design, Setting and Participants

A prospective, observational cohort study of all cases of MRSA bacteraemia in England from 1 October 2012 to 30 September 2013 was conducted. Linked, anonymised epidemiological and microbiological data were submitted electronically to the mandatory enhanced surveillance scheme (MESS) by infection control teams in acute National Health Service Trusts, in accordance with national policy. Mandatory data variables included patient demographics, details of hospital admission, date of bacteraemia, and location of acquisition (community or hospital).

6.2.2. Isolate Collection and Laboratory Testing

During the study period, all NHS diagnostic microbiology laboratories in England were invited to submit MRSA bloodstream isolates to the Staphylococcal Reference Laboratory, PHE Colindale, for characterisation. Isolates were cultured on nutrient agar and underwent s*pa* typing,[172] and PCR to confirm species identification and determination of the *mecA/C* and PVL status.[173] Isolates were stored at -80°C using Microbank™cryovials (Pro-Lab Diagnostics, Cheshire, UK) pending further analyses.

A total of 559 MRSA bloodstream isolates were received. Following quality control procedures 134 isolates were excluded, and 425 isolates were included in the analysis. The reasons of exclusion were as follows: duplicate isolates (n=50); not MRSA (n=15); inadequate isolate growth (n=2); isolates collected outside of the study dates (n=16); isolates submitted in error (n=3); non-bloodstream isolates (n=2); isolates from Wales (n=28); and isolates from Northern Ireland (n=18).

6.2.3 DNA Extraction and Whole-genome Sequencing

Isolates were retrieved from storage, sub-cultured onto nutrient agar slopes, and transferred to the Department of Medicine at the University of Cambridge. Each sample was cultured onto Columbia Blood Agar (Oxoid, Basingstoke, UK) and identified using a commercial latex agglutination kit (Pastorex Staph Plus, Bio Rad Laboratories, Hemel Hempstead, UK). Antimicrobial susceptibility testing was performed using the VITEK2 instrument (bioMérieux, Marcy l'Etoile, France). DNA was extracted, libraries prepared, and 150-bp paired-end sequences determined on an Illumina HiSeq2000 as described in Chapter 2 (Materials and Methods).

6.2.4 Integration of Datasets

PHE Staphylococcal reference laboratory test results were initially linked with demographic, clinical and geographic information from the MESS, and then anonymised prior to linkage to DNA sequence data by staff at the University of Cambridge.

6.2.5 Genomic Analysis

Genomes were assembled using an assembly and improvement pipeline as described in Chapter 2 (Materials and Methods). MLST STs and CCs were assigned from the sequence data as described in Chapter 2 (Materials and Methods). Sequence data were mapped using SMALT to the reference genome for each CC with more than 20 isolates in the CC group (CC5, N315, GenBank accession number BA000018; CC8, FPR3757, GenBank accession

number CP000255; CC22, EMRSA15, GenBank accession number HE681097). The core genome alignment excluding mobile genetic elements, indels and repetitive regions was generated for each clonal complex and was used in phylogenetic estimates using RAxML with 100 bootstraps.[174] Phylogenetic trees were visualised as described in Chapter 2 (Materials and Methods).

Isolates were *spa* typed using *in-silico* PCR to extract the *spa* gene X region from assembled genomes using previously described primers.[151] The *spa*-type was determined using an online *spa*-typer tool (http://spatyper.fortinbras.us/). The types generated through *in silico* s*pa*-type and laboratory determined *spa*-typing methods were compared to determine concordance.

Bacterial DNA sequences were deposited in the European Nucleotide Archive (ENA), https://www.ebi.ac.uk/ena, under study number ERP005128. Accession numbers, details of reads, depth of coverage and N50 are provided in Appendix 6.1. For subsequent analyses MRSA sequence data was sourced from the ENA from previously published studies. These included: i) The year-long Cambridgeshire prospective MRSA cohort study, including MRSA carriage and clinical isolates from Cambridgeshire in 2012-2013;[145] ii) MRSA bloodstream isolates collected by the British Society of Antimicrobial Chemotherapy (BSAC) Bacteraemia Surveillance Programme between 2001 and 2010;[76] iii) USA-300 isolates collected in New York, United States of America between 2009 and 2011;[60] iv) MRSA isolates from outbreak investigations at Cambridgeshire hospitals.[112,114,129]

6.3. Results

A total of 903 MRSA bacteraemia cases were reported to MESS at the time of extract (Figure 6.1). Gender was recorded for 98% of cases, and 584 (65%) of cases were male. Age was recorded for all but two cases, with a median age of 72 years (range 0 to 103 years; interquartile range (IQR) 56 to 84 years). A total of 111 laboratories participated in the study. Forty-seven per cent (425/903) of reported MRSA bacteraemia cases occurring in England during the study period had isolates that were sequenced and analysed (Figure 6.1). All of the 425 sequenced isolates were *mecA* positive by laboratory-PCR. PCR testing identified 8.7% (n=37) of the isolates as PVL-positive. Based on sequence data, 65% (n=276) were assigned to clonal complex 22 (CC22). Other CCs were represented at lower frequencies: CC5 n=42; CC30 n=33; CC8 n=22; CC1 n=19; CC59 n=9; CC45 n=7; other/unknown CCs n=17. The number of isolates and variation in the CCs isolated from each region is shown in Figure 6.2. No associations were found between particular CCs and community versus hospital onset (Appendix 6.2).
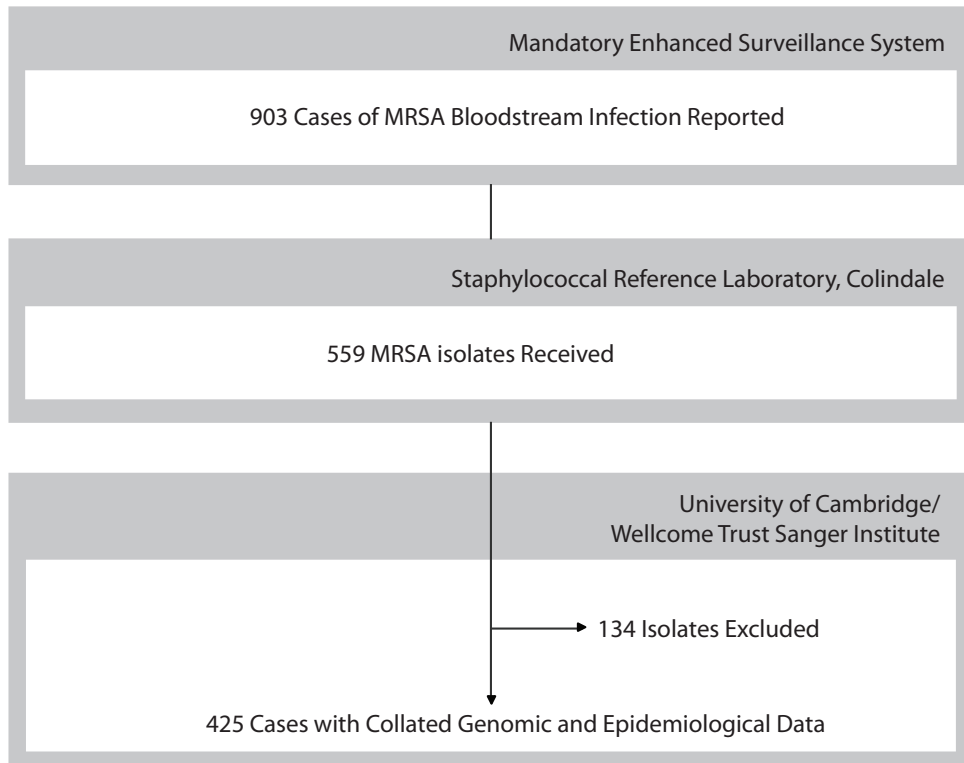
Figure 6.1. Flowchart of sample processing and analysis.
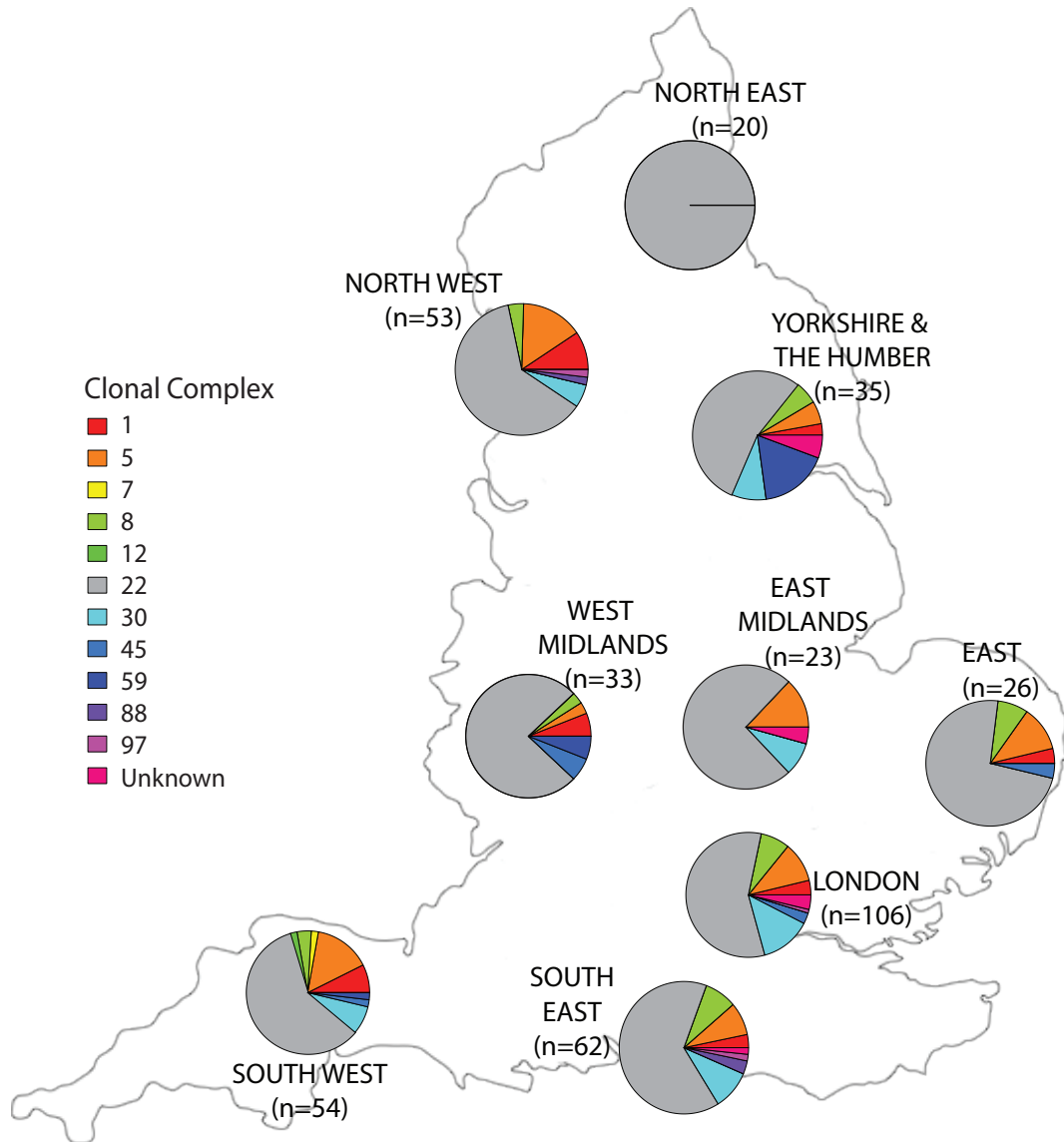
Figure 6.2. Map of England with breakdown of the proportions of each CC within the sequenced PHE bacteraemia isolates from submitting regions. The number of isolates received from each of the labelled regions is indicated in brackets. 13 isolates did not have a region assigned.

6.3.1. Comparison of Bacteraemia Surveillance and Universal MRSA Sampling

The most common clone in our collection, CC22 (n=276), was compared with CC22 genomes generated by a prospective study that sequenced MRSA isolates from every positive case (carriage and clinical samples) identified at a single diagnostic microbiology laboratory that processed samples from three hospitals and 75 GP surgeries in Cambridgeshire between April 2012 and April 2013 (Appendix 2.1). This Cambridgeshire collection was used to represent the diversity of carriage and clinical isolates within a defined geographical area, as a national collection of carriage and clinical isolates was not available. A phylogeny was constructed for the genomes from the national bacteraemia collection within this Cambridgeshire collection (Figure 6.3), in order to determine whether those isolates causing bacteraemia were clonally related, or distributed throughout the phylogenetic tree.
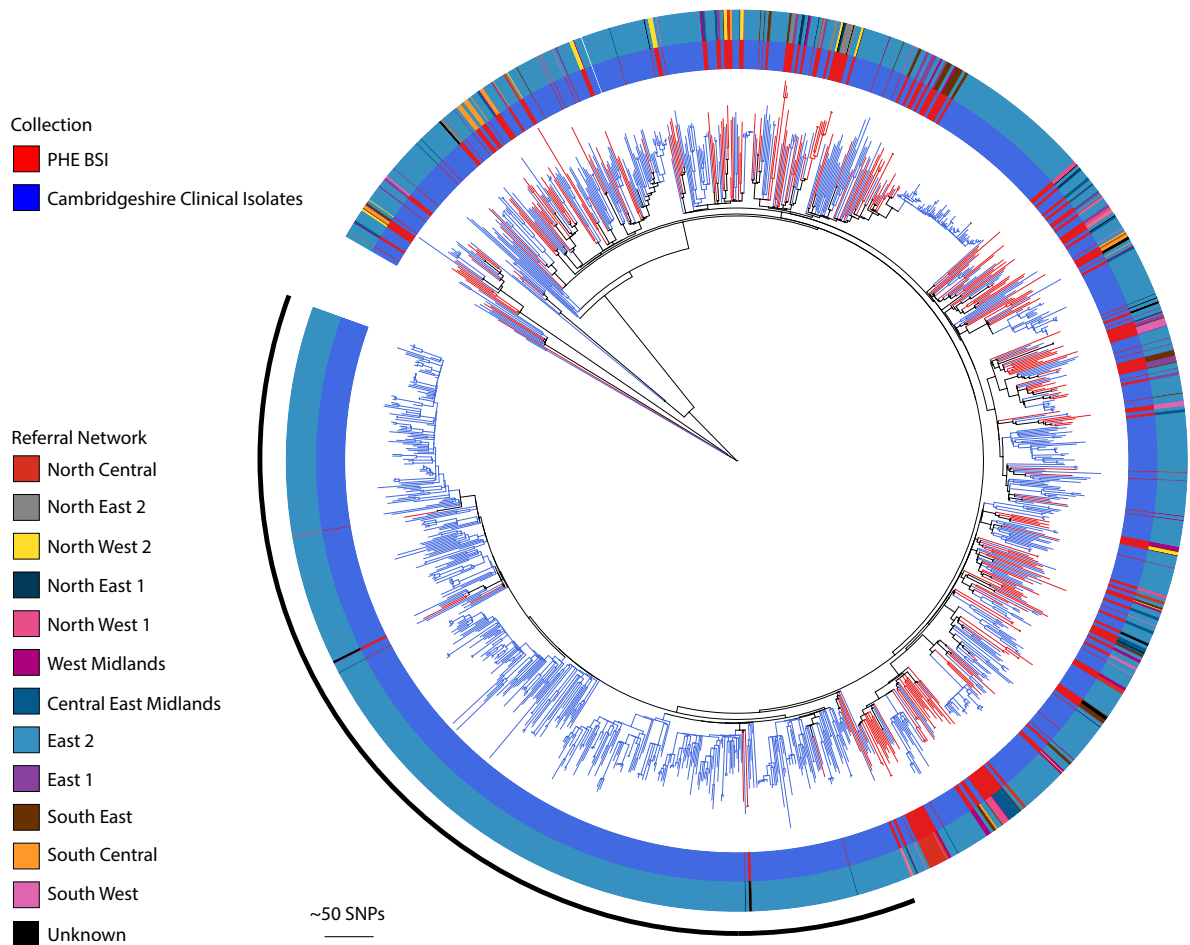
Figure 6.3. Comparison of PHE bacteraemia CC22 isolates (n=276, red branches) to the first isolate from each patient from the previously published universal sample collection from Cambridgeshire (n=1035, blue branches).[145] Mid-point rooted maximum likelihood tree based on single nucleotide polymorphisms in the core genome alignment generated after mapping against the reference genome EMRSA-15. The inner ring denotes the collection, and the outer represents referral network of submitting laboratory. The arc indicates a large clonal expansion in the Cambridgeshire region, which is underrepresented in the bacteraemia only surveillance.

As shown in Figure 6.3, isolates from our national MRSA bacteraemia collection were dispersed throughout the Cambridgeshire phylogeny, ruling out any association between a particular lineage and bacteraemia. Comparing the national bacteraemia collection to WGS of universal sampling in Cambridgeshire also demonstrates that some lineages are under-represented when undertaking bacteraemia based (rather than clinical/carriage based)

surveillance. For example, a large clonal expansion (indicated with an arc on the figure) was seen in the Cambridgeshire phylogeny, with only eight of the Cambridgeshire isolates within the national MRSA bacteraemia collection having been received from the East of England region.

To explore the effect of different sampling strategies on identified MRSA lineage diversity a comparison of CCs within three different MRSA collections was undertaken: this national MRSA bacteraemia collection (2012-2013), isolates from the Cambridgeshire study (2012-2013),[145] and MRSA bacteraemia isolates from the BSAC bacteraemia Surveillance Programme from 2000-2010 (Figure 6.4).[76] Despite the different sampling strategies and time frames, it was found that CC22 was the dominant lineage in all collections. Both of the bacteraemia-based collections showed a lower diversity of lineages than seen in the one-year Cambridgeshire study. Furthermore, the BSAC collection, which collected bacteraemia from up to 40 laboratories in the UK between 2001 and 2010, showed the most limited diversity. This may have resulted from a decline in certain lineages e.g. EMRSA-16 (CC30) during the 10-year collection period.
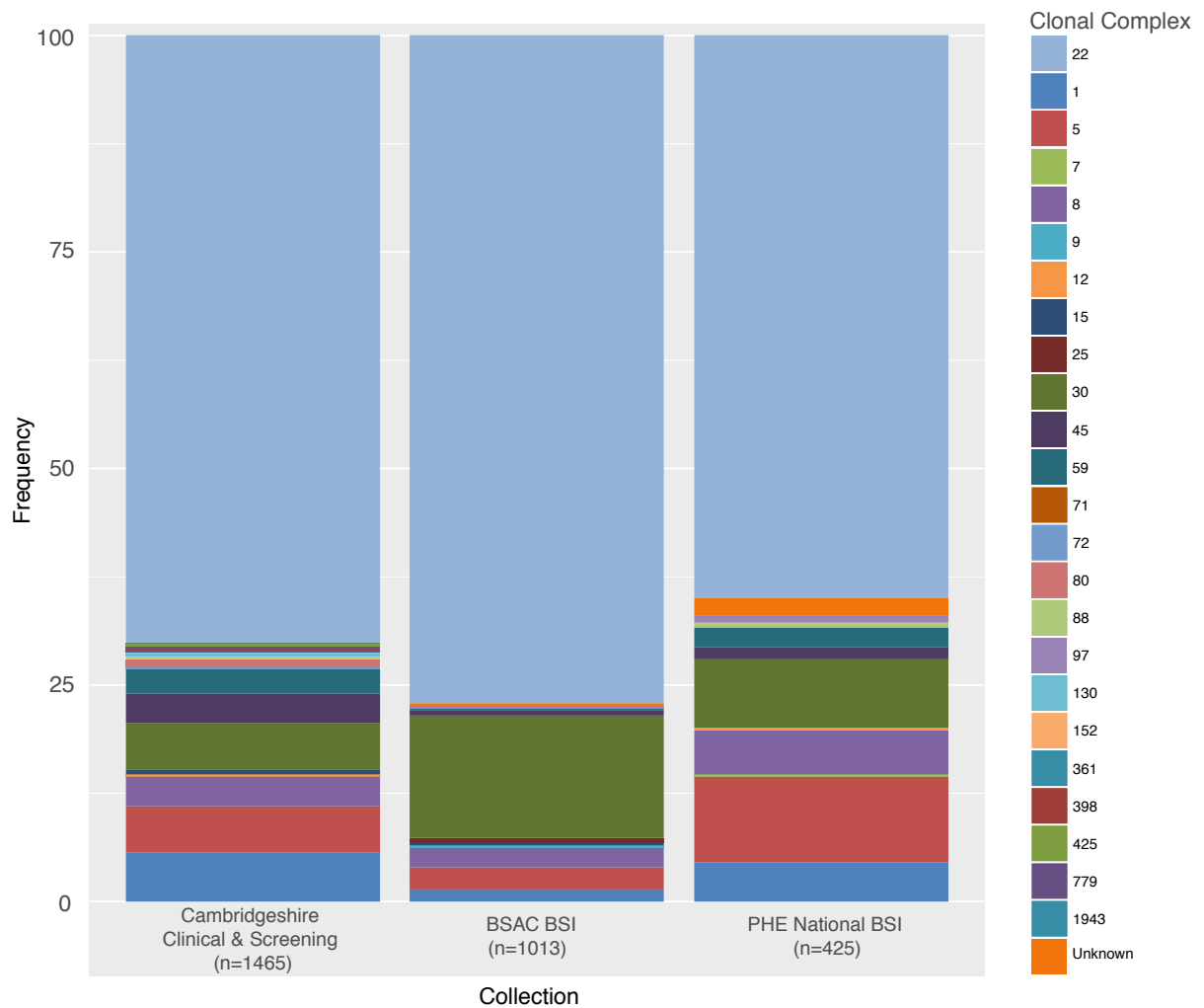
Figure 6.4. Diversity of lineages (CCs) within four collections: Carriage and clinical samples from a year-long Cambridgeshire study of MRSA;[145] the national PHE bacteraemia collection and a national BSAC bacteraemia collection.[76] Colours represent clonal complexes

6.3.2. Contextualisation of Previously Recognised Outbreaks

Reuter *et al.* previously demonstrated that it is possible to use sequence data from the BSAC MRSA bacteraemia collection (2001 to 2010) to provide genomic context for local MRSA outbreaks within a single hospital setting.[76] A similar analysis was conducted here using the national MRSA bacteraemia collection as context, to determine whether this might be

feasible using a smaller sample of bacteraemia collected during over the shorter time period

of one year. It was found that previous outbreaks in a neonatal intensive care unit,[112] and a

paediatric intensive care unit,[114] were easily identifiable as discrete clusters, as shown in

Figure 6.5. Furthermore, MRSA isolates from a suspected outbreak on a hepatology

ward,[129] were scattered throughout the phylogeny, refuting the outbreak.
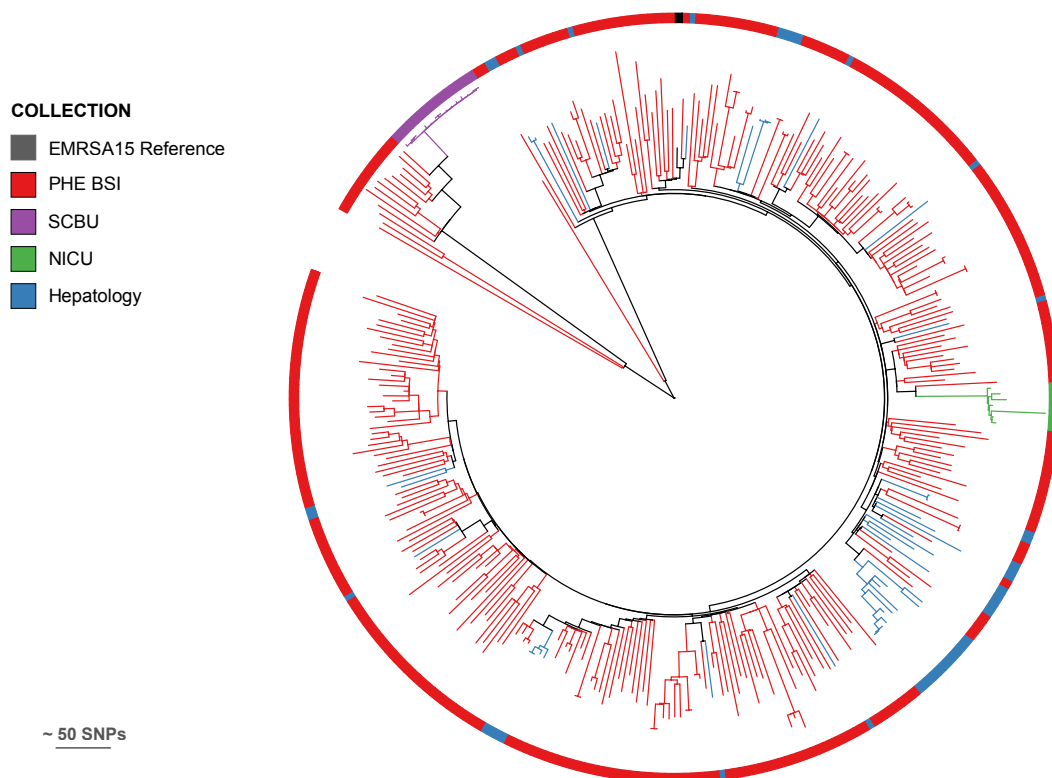
Figure 6.5. Mid-point rooted, maximum-likelihood phylogenetic tree based on SNPs in the core genome providing contextualisation of previously published outbreaks at Cambridge University Hospitals: Isolates from an outbreak on a neonatal intensive care unit (n=7, green),[112] isolates from an outbreak on a special care baby unit which extended into epidemiologically linked cases in the community (n=15, purple)[114] and isolates from a suspected but disproven outbreak on a hepatology ward (n=42, blue).[129]

6.3.4. Monitoring and Detection of Emerging or High-risk Lineages

One key aim of a national MRSA surveillance is the identification and monitoring of emerging and/or high-risk MRSA lineages. One such lineage is the USA300 lineage, which was first identified in 1999, and has subsequently caused an epidemic of SSTI in the US.[56,57] The widespread dissemination of USA300 in otherwise healthy people, and its spread into hospitals has made this a high-risk strain. However, despite multiple introductions into a number of countries, genomic surveillance has shown that to date, minimal transmission of USA300 has occurred in Europe.[66-69,175] The national MRSA bacteraemia collection was examined and it was found that six of the 22 CC8 isolates were phylogenetically defined as USA300 and were widely dispersed throughout the collection, indicating multiple introductions of USA300 into England (Figure 6.6). Given the observation that USA300 is commonly associated with SSTI (which are rarely sampled), and the limitations of bacteraemia-based sampling, it is likely that, as shown by the work presented in Chapter 4 (Regional surveillance for high-risk MRSA lineages), the prevalence of USA300 in the UK may be higher than detected here,.
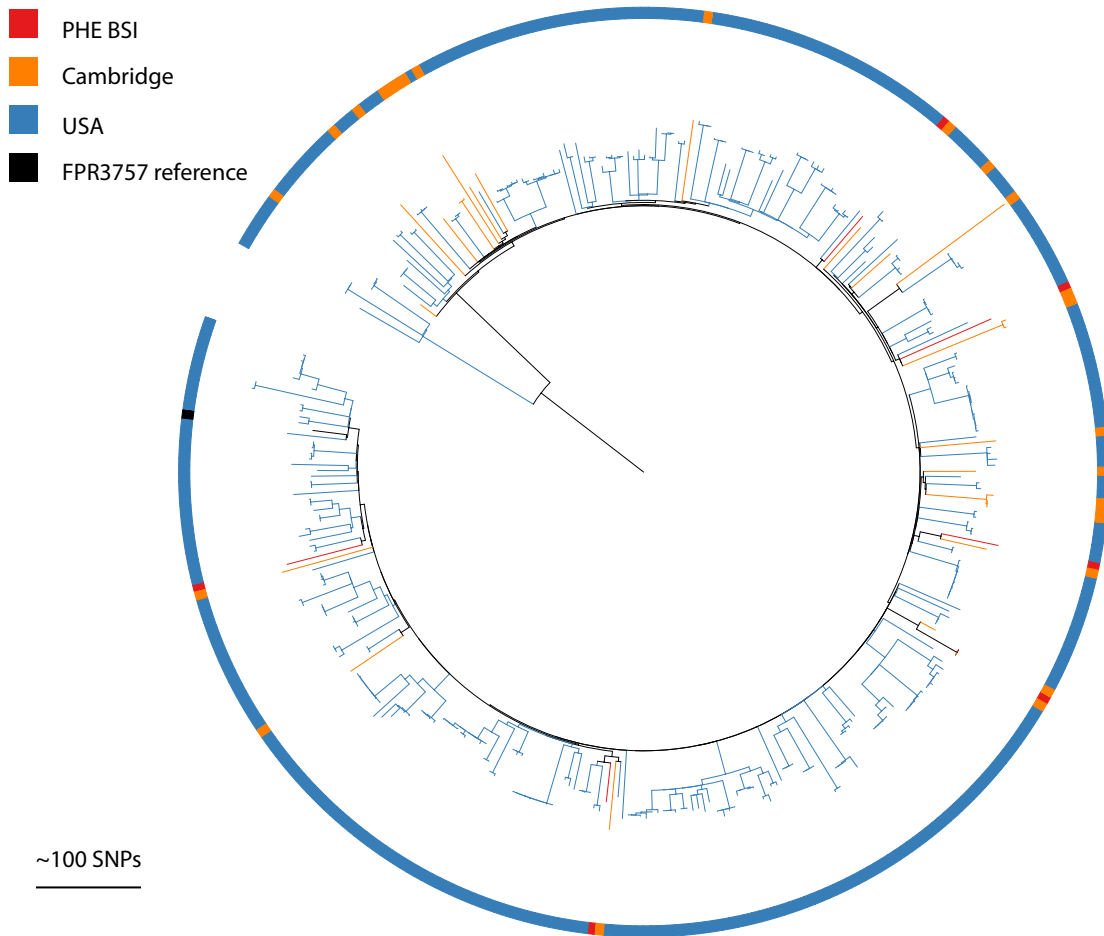
Figure 6.6. Comparison of the PHE bacteraemia USA300 isolates (n=6, red) to previously published USA300 isolates from a universal sample collection in Cambridgeshire (n=24, orange) [145] and from the US (n=348, blue) [60]. Mid-point rooted maximum likelihood tree based on SNPs in the core genome alignment generated after mapping against the reference genome FPR3757 (black).

Another potential benefit of having access to national surveillance data is the ability to identify and explore changes in molecular epidemiology on a local scale. By way of an example, it was found that an expansion of CC5 in the South West region of England (Figure 6.7), was genetically distinct from a CC5 expansion in Wales identified in the BSAC collection,[76] despite their close geographic proximity.
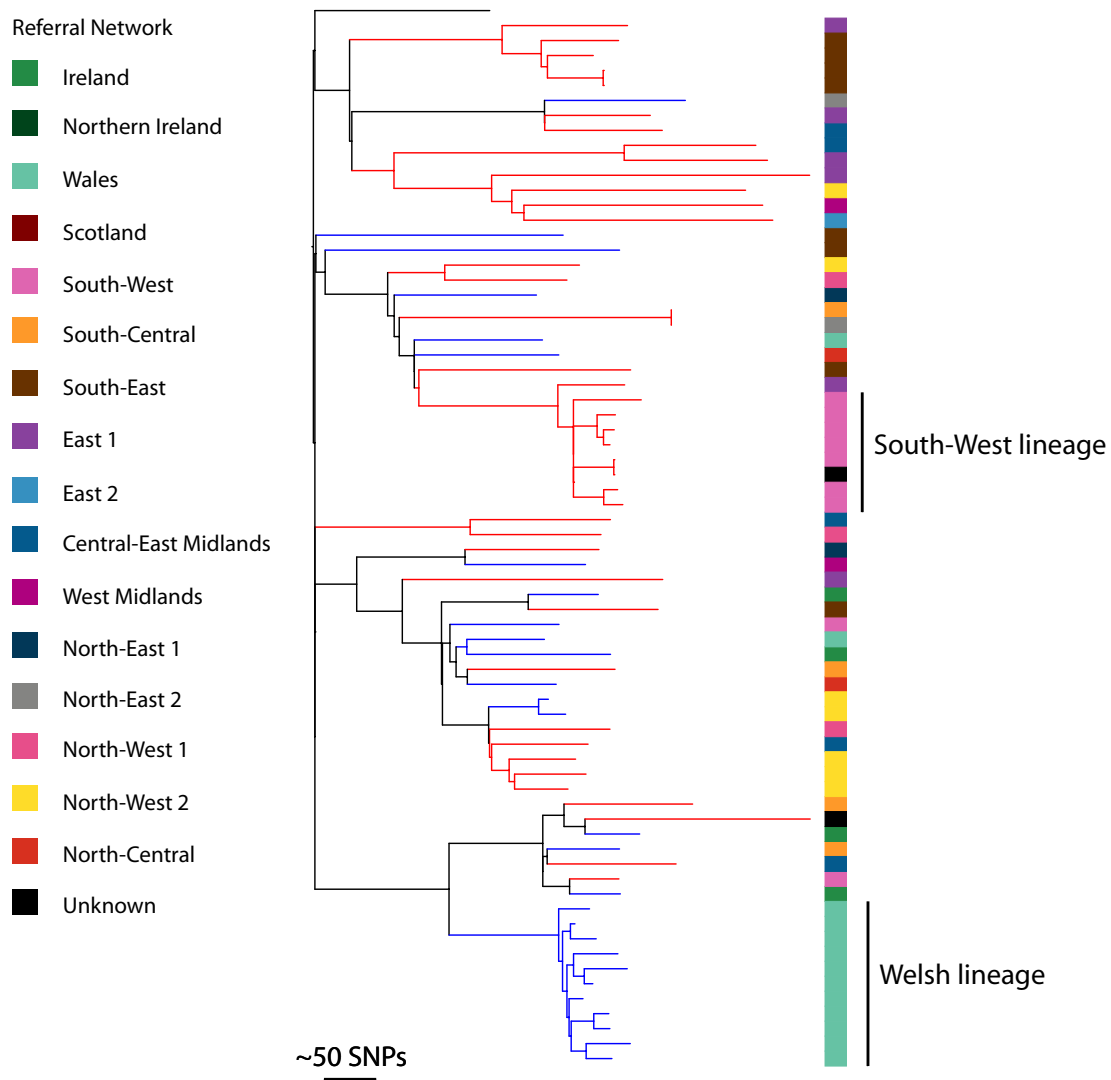
Figure 6.7. Comparison of PHE bacteraemia CC5 isolates (n=42, red branches) to those from the previously published national bacteraemia BSAC collection (BSAC, n=28, blue branches)[76]. Mid-point rooted maximum likelihood tree based on single nucleotide polymorphisms in the core genome alignment generated after mapping against the CC5 reference genome N315. Coloured bar represents referral network of submitting laboratory. Expansions within the South West of England and Welsh regions highlighted.

6.3.5. Backward Compatibility of Typing Methods

Globally, PCR determination of *spa*-type is a commonly used typing method. However, as some laboratories transition to WGS-based typing, it is important that typing methods

remain compatible. The concordance between these two methods in the national MRSA bacteraemia collection was examined. Of the 425 isolates a 98.4% concordance rate was found (Appendix 6.2), comparable to previous studies.[176,177] Of the seven isolates with discordant results, there were deletions/rearrangements within the *spa* gene of the short-read assemblies that resulted in loss of sequence complementary to the forward primer, and thus failure to amplify (*in silico*) the gene region targeted by *in silico spa*-typing.

6.4. Discussion

Mandatory enhanced surveillance for MRSA bacteraemia in England has provided in-depth information on the national decline of MRSA bacteraemia, and the changes in patient-level epidemiology that have accompanied it. However, without characterisation of systematically collected isolates, bacterial molecular epidemiology cannot be studied. This study aimed to investigate whether it was feasible to undertake combined epidemiological and genomic surveillance of MRSA bloodstream infections in England in order to address this issue.

Some challenges were encountered including obtaining bloodstream isolates from all participating hospitals (as submission was voluntary), and integrating two datasets collected through different methods (epidemiological data collected through an online database submission and isolates sent via post/courier). Despite this, the feasibility of this approach was demonstrated. The known population structure and diversity of MRSA in England was constructed, even with an incomplete collection of bloodstream isolates collected over a one-year period. A greater diversity of clones was found than that seen in a ten-year national collection of MRSA bloodstream isolates (BSAC collection) with a limited sampling strategy, but less diversity than that seen in a one-year regional collection

of carriage and clinical isolates (Cambridgeshire study). A reasonable first step in MRSA surveillance is to assess existing genomic diversity,[171] and our study demonstrates that this can be achieved and could feasibly be extended over time to generate a comprehensive national genomic database to monitor changes in clonal diversity.

Prior to April 2017, any MRSA bacteraemia isolates submitted to PHE were routinely characterised by *spa*-typing,[172] and PCR to confirm species identification alongside determination of *mecA/C* and PVL status.[173] As typing methods evolve and WGS becomes increasingly routine, backward compatibility with previous methods ensures the continued utility of typed historical collections. Laboratory *spa*-typing and *in silico spa*-typing from short read WGS data have been shown to be largely comparable in a limited number of studies,[176,177] despite the high density of repeats within the *spa* gene region. The study showed over 98% concordance between laboratory and genomic *spa*-typing methods which, reassuringly, confirms compatibility with historical data.

A further potential benefit of prospective sequencing of MRSA bloodstream isolates and a centralised national database is the ability to provide genomic context to confirm or refute outbreaks on a local or a national scale. This would be an invaluable resource as long as there is open access to anonymised (non-identifiable) data and to bioinformatics tools to analyse them rapidly and easily. The challenges will be to ensure standardised methods, development of strategies to avoid duplication of samples, and establishment of a large, comprehensive, open access, anonymised database where data could be deposited, curated, and accessible for public health benefit. As discussed previously, web-based, open-access software packages that are potentially suitable for this purpose are already being developed.[135,136] Apart from the ability to detect emerging or potentially high-risk MRSA

clones retrospectively, on-going sampling and analysis will enable detection in real-time. In this study it was found that that the high-risk USA300 lineage, an epidemic cause of SSTI in the USA, has spread to the UK and is causing bloodstream infections across England. The genomic data suggest multiple introductions of USA300, supporting the regional findings of Chapter 4 (Regional surveillance for high-risk MRSA lineages). However, the use of bacteraemia rather than clinical isolate-based surveillance limits ability to analyse this further. However, using the PHE bacteraemia collection it was possible to identify a local expansion of CC5 causing bacteraemia in the South West region of England, where local investigations suggest this clone has been problematic.[178] Thus timely, routine WGS of PHE bacteraemia isolates combined with local epidemiological data could potentially identify novel and/or pathogenic lineages in real time, and could be used to trigger local/regional investigations and interventions.

This study had several limitations. The systems for collecting epidemiological data and bacterial isolates were separate and different, leading to high rates of sample exclusion. This challenge of capturing and integrating both types of data could be overcome in practice by submitting epidemiological and laboratory data to a single data collection system. Submission of bloodstream isolates was voluntary, with many reported cases having no corresponding isolate referred for characterisation; this may have introduced bias into the analysis. This could be addressed by having mandatory submission of isolates for all reported cases. Finally, a cost/benefit analysis of this approach was not conducted. Despite these limitations, it has been demonstrated that prospective epidemiological and genomic surveillance of MRSA bloodstream infections is feasible, has numerous potential benefits, and could provide a valuable public health resource in England and beyond.

Chapter 7. Conclusions

7.1 Conclusions and Future Directions

MRSA is classified as 'high priority' in the WHO's global priority list for antibiotic resistant bacteria.[179] Furthermore, as the threat from other antibiotic resistant bacteria increases, MRSA has long been considered the prototype of antimicrobial resistant pathogens,[180] and experience in this field can be applied to other antibiotic resistant species. Surveillance plays a major part in the prevention and control of infectious diseases, and the development of technologies in the last 20 years has provided opportunities to optimise this. In her 2011 annual report, the UK Chief Medical Officer discussed the significant opportunity offered by novel technologies such as whole-genome sequencing to undertake, 'more rapid and more discriminating surveillance'. It highlighted the importance of any national strategy to implement genomics within diagnostic services in including, 'the delivery of real surveillance data as an integral part'.[130]

One of the primary aims of surveillance within a healthcare facility is to detect disease transmission and outbreaks, a key role of an infection control team. Despite being been shown as highly effective in providing the additional resolution required for effective investigation of MRSA outbreaks, WGS is not yet routinely used in day-to-day MRSA infection control practice in England. Instead, because they are routinely available, ASPs and epidemiology are typically used to define potential outbreaks and to direct where further typing should be utilised. Chapter 3 (Evaluation of ASPs to detect MRSA transmission) described the first study to systematically analyse the utility of ASP and epidemiology in determining putative MRSA transmission events. This demonstrated that

the performance of ASP and epidemiology was poor, but improved slightly when isolates with the dominant ASP in the collections were excluded. It also performed better when individuals shared a residential post code rather than a stay in the same ward at the same time, but there was little variation when *spa*-typing was included.

The persistence of ASPs as surrogates for determining transmission events is largely due to their widespread availability, as the processing of each organism to guide antibiotic therapy for clinicians produces the ASP as a (potentially) helpful by-product. WGS offers the potential to both identify organisms and determine the presence/absence of drug resistance mutations and genes. Because of variation in gene expression, genotypic resistance does not always cause phenotypic resistance. Therefore, a high level of concordance between phenotypic (current practice) and genotypic susceptibility results is required for this to be accepted into routine practice. *S. aureus* has been studied extensively in this regard,[139,140] and Gordon *et al.*, 2014, have shown that WGS is as sensitive and specific as routine antimicrobial susceptibility testing methods.[138] If WGS was transitioned into routine susceptibility testing, then the use of WGS data instead of ASPs for detecting transmission as part of infection control practice would be inevitable. In the meantime, future work should combine large scale, systematic, real-time WGS of MRSA alongside ASPs and epidemiology to determine how WGS can be best targeted to improve outbreak detection, prior to the potential future integration of WGS into routine practice.

Chapter 4 (Regional surveillance for high-risk MRSA lineages) explored the use of systematic surveillance to investigate potentially high-risk MRSA lineages. It determined the prevalence of USA300 MRSA in Cambridgeshire, providing the first insights into the burden of USA300 in an area of the UK, with phylogenetic analysis revealing multiple

introductions and household transmission. This supports the findings from other studies elsewhere in Europe, such as France and Switzerland.[72,181] In contrast to these studies from elsewhere in continental Europe, no isolates of USA300-LV from Latin America were identified in the sampled Cambridgeshire population. Whilst USA300 is considered a high-risk lineage globally because of its history of rapid spread in the US,[171] ST2371 MRSA is an example of a lineage which has primarily proven problematic in the Cambridgeshire region in England. Despite an outbreak of this lineage being declared closed, this work uncovered ten isolates which were closely related to the original outbreak isolates both from a genomic and epidemiological perspective. Although the ward outbreak had indeed been brought to a close, with no systematic surveillance programme monitoring the incidence of non-invasive MRSA infections, this lineage continued to cause disease in a group of linked individuals in the community, highlighting limitations in the current strategies to monitor cases.

Chapter 4 (Regional surveillance for high-risk MRSA lineages) therefore shows how WGS could assist in the surveillance and monitoring of such high-risk lineages, and more broadly, it indicates that systematic WGS within a sentinel laboratory could be utilised as an MRSA surveillance mechanism. Future work should explore the practicalities and costings of different strategies for applying genomic surveillance to potentially high-risk lineages, including sentinel surveillance within specific laboratories as part of national surveillance. Through modelling different surveillance strategies, the optimal strategy can be determined and importantly, costed.

Chapter 5 (Detection of cryptic local MRSA transmission) used systematic surveillance and WGS to identify transmission of the typically nosocomial lineage, ST22, within a

Cambridgeshire community. Based on the higher overall prevalence of MRSA in hospitals versus the community in the last few decades it has been assumed that the predominant directionality of spread of ST22, the most common MRSA lineage in England, is from hospitals into the community. The work in Chapter 5 (Detection of cryptic local MRSA transmission) challenges this view and demonstrates the need to consider GP surgeries as transmission hotspots. Previous studies conducted in the UK have isolated ST22 from the community,[168,169,182] but bacterial typing lacked sufficient resolution to infer transmission. As one of the cases carried the same MRSA lineage for nearly three years and the diversity within the isolates from this case encompassed that of isolates from all other cases, this study also highlighted the important role of persistent MRSA carriers in the transmission of MRSA. Since writing, the important role of persistent carries in MRSA transmission has been confirmed in a comprehensive genomic study by Gordon *et al.*, 2017.[116]

Further work would aim to determine those factors which contributed to the success of this particular ST22 MRSA lineage in transmitting within the community. This would require a comprehensive analysis of patient, bacterial and environmental factors. However, the historical nature of the outbreak would be a limitation to the analysis. From a wider perspective, the role of infection prevention and control in the community will become increasingly relevant as initiatives are rolled out that increase delivery of care outside of hospitals,[108] meaning altered demands on infection services.[168] The implementation of surveillance WGS to control procedures may be a necessary tool if MRSA transmission is to be targeted by rapid interventions.

Through exploring the integration of epidemiological and genomic surveillance at a national level, the investigation described in Chapter 6 (National surveillance of MRSA

bacteraemia) demonstrated that it is feasible to undertake combined epidemiological and genomic surveillance of MRSA bacteraemia in England. Logistical challenges were encountered which could help inform practice for the integration of these services in PHE in future. Challenges included obtaining bloodstream isolates from all participating hospitals (as submission was voluntary), and integrating two datasets collected through different methods (epidemiological data collected through an online database submission and isolates sent via post/courier).

A major advantage of sequencing MRSA isolates is the ability to share and collate genome sequence data to build up national and international databases. A number of bacteraemia surveillance systems already exist e.g. the English mandatory enhanced surveillance system, the voluntary British Society of Antimicrobial Chemotherapy bacteraemia Surveillance Programme, and the voluntary European Antimicrobial Resistance Surveillance Network. Whilst each system has different aims and objectives, sampling criteria and data collection methods, the digital interchangeability of sequence data creates an opportunity to collaborate and share genome sequence data whilst producing a sustainable, on-going resource if the isolates were sequenced.

7.2. Closing Remarks

This thesis has used phenotypic, epidemiological and genomic data to investigate MRSA surveillance at the regional and national level. It has quantified the sensitivity and specificity of ASPs and epidemiology versus the gold standard of WGS and epidemiology in determining MRSA transmission. It has also revealed how WGS can be used alongside systematic surveillance to investigate potentially high-risk lineages of MRSA. It has shown

that WGS can feasibly be integrated into an existing national surveillance programme and

can provide information to enhance surveillance at a national and regional level.

Chapter 8. References

1.   Cowan ST and Steel KJ. Manual for the Identification of Medical Bacteria. Cambridge University Press. pp 53.

2.   Wertheim HF, Melles DC, Vos MC, et al.. The role of nasal carriage in *Staphylococcus aureus* infections. Lancet Infectious Diseases 2005;5:751-62.

3.   Kluytmans J, van Belkum A, Verbrugh H. Nasal carriage of *Staphylococcus aureus*: epidemiology, underlying mechanisms, and associated risks. Clinical Microbiology Reviews 1997;10:505-20.

4.   Lowy FD. *Staphylococcus aureus* infections. New England Journal of Medicine 1998; 339: 520-5321998.

5.   *Staphylococcus aureus* (including toxic shock syndrome). In: Mandell GL, Bennett JE, Dolin R, eds. Mandell, Douglas and Bennett's Principles and Practice of Infectious Diseases. 4th ed. Vol. 2. New York: Churchill Livingstone, 2017:1754-77.2017.

6.   von Eiff C, Becker K, Machka K, et al. Nasal carriage as a source of *Staphylococcus aureus* bacteremia. New England Journal of Medicine 2001;344:11-6.

7.   Wertheim HF, Vos MC, Ott A, et al. Risk and outcome of nosocomial *Staphylococcus aureus* bacteraemia in nasal carriers versus non-carriers. Lancet 2004;364:703-5.

8.   Enright MC, Robinson DA, Randle G, et al. The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA). Proceedings of the National Academy of Sciences 2002;99:7687-92.

9.   Katayama Y, Ito T, Hiramatsu K. A New Class of Genetic Element, *Staphylococcus* Cassette Chromosome *mec*, Encodes Methicillin Resistance in *Staphylococcus aureus*. Antimicrobial Agents and Chemotherapy 2000;44:1549-55.

10.   International Working Group on the Classification of Staphylococcal Casette Chromosome Elements (IWG-SCC). Classification of Staphylococcal Cassette Chromosome *mec* (SCC*mec*): Guidelines for Reporting Novel SCC*mec* Elements. Antimicrobial Agents and Chemotherapy 2009;53:4961-7.

11.   Jevons MP. "Celbenin"-resistant Staphylococci. British Medical Journal 1961;1:124-5.

12.   Harkins CP, Pichon B, Doumith M, et al. Methicillin-resistant *Staphylococcus aureus* emerged long before the introduction of methicillin into clinical practice. Genome Biology 2017;18:130.

13.   International Working Group on the Staphylococcal Cassette Chromosome elements (IWG-SCC) SCC*mec* up-to-date. http://www.sccmec.org/Pages/SCC_TypesEN.html Last accessed: 14th March 2018.

14.   Ubukata K, Nonoguchi R, Matsuhashi M, et al. Expression and inducibility in *Staphylococcus aureus* of the *mec*A gene, which encodes a methicillin-resistant *S. aureus*-specific penicillin-binding protein. Journal of Bacteriology 1989;171:2882-5.

15.   Hartman B, Tomasz A. Altered penicillin-binding proteins in methicillin-resistant strains of *Staphylococcus aureus*. Antimicrobial agents and Chemotherapy 1981;19:726-35.

16.   Ba X, Harrison EM, Edwards GF, et al. Novel mutations in penicillin-binding protein genes in clinical *Staphylococcus aureus* isolates that are methicillin-resistant on susceptibility testing, but lack the *mec* gene. Journal of Antimicrobial Chemotherapy 2014;69:594-7.

17.	Garcia-Alvarez L, Holden MT, Lindsay H, et al. Meticillin-resistant *Staphylococcus aureus* with a novel *mec*A homologue in human and bovine populations in the UK and Denmark: a descriptive study. Lancet Infectious Diseases 2011;11:595-603.

18.	Becker K, van Alen S, Idelevich EA, et al. Plasmid-Encoded Transferable *mec*B-Mediated Methicillin Resistance in *Staphylococcus aureus*. Emerging Infectious Diseases 2018;24:242-8.

19.	Zhanel GG, Sniezek G, Schweizer F, et al. Ceftaroline: a novel broad-spectrum cephalosporin with activity against meticillin-resistant *Staphylococcus aureus*. Drugs 2009;69:809-31.

20.	Page MG. Ceftobiprole - a case study. Expert opinion on Drug Discovery 2007;2:115-29.

21.	Cadena J, Thinwa J, Walter EA, et al. Risk factors for the development of active methicillin-resistant *Staphylococcus aureus* (MRSA) infection in patients colonized with MRSA at hospital admission. American Journal of Infection Control 2016;44:1617-21.

22.	Cosgrove SE, Qi Y, Kaye KS, et al. The impact of methicillin resistance in *Staphylococcus aureus* bacteremia on patient outcomes: mortality, length of stay, and hospital charges. Infection Control and Hospital Epidemiology 2005;26:166-74.

23.	Fowler VG, Jr., Miro JM, Hoen B, et al. *Staphylococcus aureus* endocarditis: a consequence of medical progress. Journal of the American Medical Association 2005;293:3012-21.

24.	Chang FY, Peacock JE, Musher DM, et al. *Staphylococcus aureus* bacteremia: recurrence and the impact of antibiotic treatment in a prospective multicenter study. Medicine 2003;82:333-9.

25.	Stefani S, Chung DR, Lindsay JA, et al. Meticillin-resistant *Staphylococcus aureus* (MRSA): global epidemiology and harmonisation of typing methods. International Journal of Antimicrobial Agents 2012;39:273-82.

26.	Wyllie DH, Walker AS, Miller R, et al. Decline of meticillin-resistant *Staphylococcus aureus* in Oxfordshire hospitals is strain-specific and preceded infection-control intensification. BMJ open 2011;1:e000160.

27.	Chambers HF, DeLeo FR. Waves of Resistance: *Staphylococcus aureus* in the Antibiotic Era. Nature Reviews Microbiology 2009;7:629-41.

28.	Jevons MP, Coe AW, Parker MT. Methicillin resistance in Staphylococci. Lancet 1963;1:904-7.

29.	Duckworth GJ, Lothian JL, Williams JD. Methicillin-resistant *Staphylococcus aureus*: report of an outbreak in a London teaching hospital. Journal of Hospital Infection 1988;11:1-15.

30.	Reacher MH, Shah A, Livermore DM, et al. Bacteraemia and antibiotic resistance of its pathogens reported in England and Wales between 1990 and 1998: trend analysis. BMJ (Clinical research ed) 2000;320:213-6.

31.	Health Protection Agency. Results from the mandatory surveillance of MRSA bacteraemia. http://webarchive.nationalarchives.gov.uk/20140714025116tf_/http://www.hpa.org.uk/web/HPAweb&HPAwebStandard/HPAweb_C/1233906819629. Last accessed: 16th March 2019.

32.	Public Health England. Annual Epidemiological Commentary. 2016-2017. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/634675/Annual_epidemiological_commentary_2017.pdf. Last accessed: 21st Feb 2019.

33.	Richardson JF, Reith S. Characterization of a strain of methicillin-resistant *Staphylococcus aureus* (EMRSA-15) by conventional and molecular methods. The Journal of Hospital Infection;25:45-52.

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

35.   Rosendal K, Bulow P, Bentzon MW, Eriksen KR. *Staphylococcus aureus* strains isolated in Danish hospitals from January 1st, 1966, to December 31st, 1974. Acta pathologica et microbiologica Scandinavica Section B, Microbiology 1976;84b:359-68.

36.   Kayser FH. Methicillin-resistant staphylococci 1965-75. The Lancet 1975;2:650-3.

37.   Wyllie D, Paul J, Crook D. Waves of trouble: MRSA strain dynamics and assessment of the impact of infection control. The Journal of Antimicrobial Chemotherapy 2011;66:2685-8.

38.   Johnson AP, Davies J, Guy R, et al. Mandatory surveillance of methicillin-resistant *Staphylococcus aureus* (MRSA) bacteraemia in England: the first 10 years. Journal of Antimicrobial Chemotherapy 2012;67:802-9.

39.   Lawes T, Lopez-Lozano JM, Nebot C, et al. Turning the tide or riding the waves? Impacts of antibiotic stewardship and infection control on MRSA strain dynamics in a Scottish region over 16 years: non-linear time series analysis. BMJ Open 2015 26;5(3):e006596.

40.   Public Health England. UK Standards for Microbiology Investigations. Investigation of Specimens for Screening for MRSA. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/674688/B_29i6_under_review.pdf. Last accessed: 23rd April 2019.

41.   Angela Kearns. Personal communication.

42.   Maiden MC, Bygraves JA, Feil E, er al. Mutlilocus Sequence Typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. Proceedings of the National Academy of Science, USA, 1998 95:3140-3145. 1998.

43.   Maiden MC. Multilocus sequence typing of bacteria. Annual Review of Microbiology 2006;60:561-88.

44.   Enright MC, Day NP, Davies CE, et al. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. Journal of Clinical Microbiology 2000;38:1008-15.

45.   Feil EJ, Cooper JE, Grundmann H, et al. How Clonal Is *Staphylococcus aureus*? Journal of Bacteriology 2003;185:3307-16.

46.   Frenay HM, Bunschoten AE, Schouls LM, et al. Molecular typing of methicillin-resistant *Staphylococcus aureus* on the basis of protein A gene polymorphism. European Journal of Clinical Microbiology & Infectious Diseases 1996;15:60-4.

47.   Harmsen D, Claus H, Witte W, et al. Typing of Methicillin-resistant *Staphylococcus aureus* in a University Hospital Setting by Using Novel Software for *spa* Repeat Determination and Database Management. Journal of Clinical Microbiology 2003;41:5442-8.

48.   Strommenger B, Kettlitz C, Weniger T, et al. Assignment of *Staphylococcus* Isolates to Groups by s*pa*-Typing, SmaI Macrorestriction Analysis, and Multilocus Sequence Typing. Journal of Clinical Microbiology 2006;44:2533-40.

49.   Otter JA, French GL. Community-associated meticillin-resistant *Staphylococcus aureus*: the case for a genotypic definition. The Journal of Hospital infection 2012;81:143-8.

50.   Moran GJ, Krishnadasan A, Gorwitz RJ, et al. Methicillin-resistant *S. aureus* infections among patients in the emergency department. The New England Journal of Medicine 2006;355:666-74.

51.   EUCAST Disk Diffusion Method for Antimicrobial Susceptibility Testing. Version 5.0 2015, www.eucast.org. Last accessed: 16th March 2019.

52.	The European Committee on Antimicrobial Susceptibility Testing. Breakpoint tables for interpretation of MICs and zone diameters. Version 5.0 2015. http://www.eucast.org. Last accessed: 16th March 2019.

53.	Udo EE, Pearman JW, Grubb WB. Genetic analysis of community isolates of methicillin-resistant *Staphylococcus aureus* in Western Australia. The Journal of Hospital Infection 1993;25:97-108.

54.	Chambers HF. The changing epidemiology of *Staphylococcus aureus*. Emerging Infectious Diseases 2001;7:178-82.

55.	Deurenberg RH, Stobberingh EE. The molecular evolution of hospital- and community-associated methicillin-resistant *Staphylococcus aureus*. Current Molecular Medicine 2009;9:100-15.

56.	David MZ, Daum RS. Community-associated methicillin-resistant *Staphylococcus aureus*: epidemiology and clinical consequences of an emerging epidemic. Clinical Microbiology Reviews 2010;23:616-87.

57.	Centers for Disease Control and Prevention (CDC). Methicillin-resistant *Staphylococcus aureus* skin or soft tissue infections in a state prison - Mississippi, 2000. Morbidity Mortality Weekly Reports 2001:919-22.

58.	Jenkins TC, McCollister BD, Sharma R, et al. Epidemiology of healthcare-associated bloodstream infection caused by USA300 strains of methicillin-resistant *Staphylococcus aureus* in 3 affiliated hospitals. Infection Control and Hospital Epidemiology 2009;30:233-41.

59.	Seybold U, Kourbatova EV, Johnson JG, et al. Emergence of community-associated methicillin-resistant *Staphylococcus aureus* USA300 genotype as a major cause of health care-associated blood stream infections. Clinical Infectious Diseases 2006;42:647-56.

60.	Uhlemann A-C, Dordel J, Knox JR, et al. Molecular tracing of the emergence, diversification, and transmission of S. aureus sequence type 8 in a New York community. Proceedings of the National Academy of Sciences of the United States of America 2014;111:6738-43.

61.	Alam MT, Read TD, Petit RA, 3rd, et al. Transmission and microevolution of USA300 MRSA in U.S. households: evidence from whole-genome sequencing. mBio 2015;6:e00054.

62.	Planet PJ, Diaz L, Kolokotronis S-O, et al. Parallel Epidemics of Community-Associated Methicillin-resistant *Staphylococcus aureus* USA300 Infection in North and South America. Journal of Infectious Diseases 2015;212(12):1874-82.

63.	Nurjadi D, Friedrich-Janicke B, Schafer J, et al. Skin and soft tissue infections in intercontinental travellers and the import of multi-resistant *Staphylococcus aureus* to Europe. Clinical Microbiology and Infection 2015;21:567.e1-.e10.

64.	Nimmo GR. USA300 abroad: global spread of a virulent strain of community-associated methicillin-resistant *Staphylococcus aureus*. Clinical Microbiology and Infection 2012;18:725-34.

65.	Ellington MJ, Hope R, Livermore DM, et al. Decline of EMRSA-16 amongst methicillin-resistant *Staphylococcus aureus* causing bacteraemias in the UK between 2001 and 2007. The Journal of Antimicrobial Chemotherapy 2010;65:446-8.

66.	Seidl K, Leimer N, Palheiros Marques M, et al. USA300 methicillin-resistant *Staphylococcus aureus* in Zurich, Switzerland between 2001 and 2013. International Journal of Medical Microbiology 2014;304:1118-22.

67.	Baud O, Giron S, Aumeran C, et al. First outbreak of community-acquired MRSA USA300 in France: failure to suppress prolonged MRSA carriage despite decontamination procedures. European Journal of Clinical Microbiology & Infectious Diseases 2014;33:1757-62.

68.   Blanco R, Tristan A, Ezpeleta G, et al. Molecular epidemiology of Panton-Valentine leukocidin-positive *Staphylococcus aureus* in Spain: emergence of the USA300 clone in an autochthonous population. Journal of Clinical Microbiology 2011;49:433-6.

69.   van der Mee-Marquet N, Poisson DM, Lavigne JP, et al. The incidence of *Staphylococcus aureus* ST8-USA300 among French pediatric inpatients is rising. European Journal of Clinical Microbiology & Infectious Diseases 2015;34(5):935-42.

70.   Ruppitsch W, Stoger A, Schmid D, et al. Occurrence of the USA300 community-acquired *Staphylococcus aureus* clone in Austria. Eurosurveillance 2007;12:E071025 1.

71.   Rolo J, Miragaia M, Turlej-Rogacka A, et al. High Genetic Diversity among Community-Associated *Staphylococcus aureus* in Europe: Results from a Multicenter Study. PloS one 2012;7:e34768.

72.   Glaser P, Martins-Simões P, Villain A, et al. Demography and Intercontinental Spread of the USA300 Community-Acquired Methicillin-resistant *Staphylococcus aureus* Lineage. mBio 2016;7.

73.   Ellington MJ, Yearwood L, Ganner M, East C, Kearns AM. Distribution of the ACME-*arc*A gene among methicillin-resistant *Staphylococcus aureus* from England and Wales. The Journal of Antimicrobial Chemotherapy 2008;61:73-7.

74.   Patel M, Thomas HC, Room J, Wilson Y, Kearns A, Gray J. Successful control of nosocomial transmission of the USA300 clone of community-acquired meticillin-resistant *Staphylococcus aureus* in a UK paediatric burns centre. The Journal of Hospital Infection 2013;84:319-22.

75.   Otter JA, Havill NL, Boyce JM, French GL. Comparison of community-associated methicillin-resistant *Staphylococcus aureus* from teaching hospitals in London and the USA, 2004-2006: where is USA300 in the UK? European Journal of Clinical Microbiology & Infectious Diseases 2009;28:835-9.

76.   Reuter S, Torok ME, Holden MT, et al. Building a genomic framework for prospective MRSA surveillance in the United Kingdom and the Republic of Ireland. Genome Research 2016;26:263-70.

77.   Porter M, A Dictionary of Epidemiology. Oxford University Press. 6th Ed.

78.   Nsubuga P, White ME, Thacker SB, et al. Public Health Surveillance: A Tool for Targeting and Monitoring Interventions. In: Jamison DT, Breman JG, Measham AR, et al., editors. Disease Control Priorities in Developing Countries. 2nd edition. Washington (DC): World Bank; 2006. Chapter 53. Available from: http://www.ncbi.nlm.nih.gov/books/NBK11770/. Last accessed: 30th October 2018.

79.   Redmond NM, Davies R, Christensen H, et al. The TARGET cohort study protocol: a prospective primary care cohort study to derive and validate a clinical prediction rule to improve the targeting of antibiotics in children with respiratory tract illnesses. BMC Health Services Research 2013;13:322.

80.   Zygmunt FD, Dennis GC, Julie AP. Syndromic Surveillance. Emerging Infectious Diseases 2004;10:1333.

81.   World Health Organization. Immunization, Vaccines and Biologicals. Sentinel Surveillance. http://www.who.int/immunization/monitoring_surveillance/burden/vpd/surveillance_type/sentinel/en/. Last accessed: 4th June 2019.

82.   Department of Health. Sources of UK flu data: Influenza surveillance in the UK. https://www.gov.uk/guidance/sources-of-uk-flu-data-influenza-surveillance-in-the-uk. Last accessed 30th August 2018.

83.   Practical Healthcare Epidemiology. 4th Ed. Lautenbach E, Malani PN, Han JH, Shuman EK, Marschall J (Eds). Cambridge University Press. Chapter 10. Perl TM and Gase KA. Surveillance: An Overview. pp 94.

84.   Freeman R, Charlett A, Hopkins S, et al. Evaluation of a national microbiological surveillance system to inform automated outbreak detection. The Journal of Infection 2013;67:378-84.

85.   Public Health England. English Surveillance Programme for Antimicrobial Utilisation and Resistance (ESPAUR) Report 2017. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/656611/ESPAUR_report_2017.pdf. Last accessed: 16th March 2019.

86.   Public Health England. Mandatory Health Care Associated Infection Surveillance: Data quality statement. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/635429/mandatory_healthcare_associated_infection_data_quality_report.pdf. 2017.

87.   NHS England. Everyone Counts: Planning for Patients 2014/15 – 2018/19. https://www.england.nhs.uk/wp-content/uploads/2013/12/5yr-strat-plann-guid-wa.pdf Last accessed: 29/03/2018.

88.   Public Health England. Post Infection Review Guidance. https://improvement.nhs.uk/resources/mrsa-guidance-post-infection-review/ Last accessed 29/03/2018.

89.   Public Health England. Update on the reporting and monitoring arrangements and post-infection review process for MRSA bloodstream infections. March 2018. https://improvementnhsuk/resources/mrsa-guidance-post-infection-review/. Last accessed: 16th March 2019.

90.   Department of Health. Report of the Second Phase of the Review of NHS Pathology Services in England. 2008. Chaired by Lord Carter of Coles. http://webarchive.nationalarchives.gov.uk/20130124044941/http:/www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_091984.pdf. Last accessed: 16th March 2019.

91.   National Institute for Health and Care Excellence. Infection prevention and control. Quality standard [QS61, April 2014] https://www.nice.org.uk/guidance/qs61/chapter/introduction. Last accessed: 1st June 2019.

92.   Coia JE, Duckworth GJ, Edwards DI, et al. Guidelines for the control and prevention of meticillin-resistant Staphylococcus aureus (MRSA) in healthcare facilities. The Journal of Hospital Infection 2006;63 Suppl 1:S1-44.

93.   Calfee DP, Salgado CD, Milstone AM, et al. Strategies to prevent methicillin-resistant *Staphylococcus aureus* transmission and infection in acute care hospitals: 2014 update. Infection Control and Hospital Epidemiology 2014;35 Suppl 2:S108-32.

94.   Batra R, Cooper BS, Whiteley C, et al. Efficacy and limitation of a chlorhexidine-based decolonization strategy in preventing transmission of methicillin-resistant *Staphylococcus aureus* in an intensive care unit. Clinical Infectious Diseases 2010;50:210-7.

95.   Cooper BS, Stone SP, Kibbler CC, et al. Isolation measures in the hospital management of methicillin-resistant *Staphylococcus aureus* (MRSA): systematic review of the literature. BMJ 2004;329:533.

96.   Loveday HP, Pellowe CM, Jones SR, Pratt RJ. A systematic review of the evidence for interventions for the prevention and control of meticillin-resistant *Staphylococcus aureus* (1996-2004): report to the Joint MRSA Working Party (Subgroup A). The Journal of Hospital Infection 2006;63 Suppl 1:S45-70.

97.   Sandri AM, Dalarosa MG, Ruschel de Alcantara L, et al. Reduction in incidence of nosocomial methicillin-resistant *Staphylococcus aureus* (MRSA) infection in an intensive care unit: role of treatment with mupirocin ointment and chlorhexidine baths for nasal carriers of MRSA. Infection Control and Hospital Epidemiology 2006;27:185-7.

98.   Simor AE, Phillips E, McGeer A, et al. Randomized controlled trial of chlorhexidine gluconate for washing, intranasal mupirocin, and rifampin and doxycycline versus no treatment for the eradication of methicillin-resistant *Staphylococcus aureus* colonization. Clinical Infectious Diseases 2007;44:178-85.

99.   Robotham JV, Deeny SR, Fuller C, Hopkins S, Cookson B, Stone S. Cost-effectiveness of national mandatory screening of all admissions to English National Health Service hospitals for meticillin-resistant *Staphylococcus aureus*: a mathematical modelling study. The Lancet Infectious Diseases 2016;16:348-56.

100.   Department of Health. Implementation of modified admission MRSA screening guidance   for   the   NHS,   2014. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/345144/Implementation_of_modified_admission_MRSA_screening_guidance_for_NHS.pdf.)   Last accessed: 16th March 2019.

101.   Roth JA, Hornung-Winter C, Radicke I, et al. Direct Costs of a Contact Isolation Day: A Prospective Cost Analysis at a Swiss University Hospital. Infection Control & Hospital Epidemiology 2017;39:101-3.

102.   Cairns S. Packer S, Reilly J, et al. Targeted MRSA screening can be as effective as universal screening. BMJ 2014; 349 :g5075

103.   Public Health England. MRSA screening and suppression in primary care guidance. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/613791/MRSA_screening_and_supression_primary_care_guidance.pdf)   Last accessed: 16th March 2019.

104.   Health and Social Care Act 2008: Code of Practice on the prevention and control of   infections   and   related   guidance. (https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/449049/Code_of_practice_280715_acc.pdf). Last accessed: 16th March 2019.

105.   Care Quality Commission. Scope of Registration Guidance. 2015. (http://www.cqc.org.uk/sites/default/files/20151230_100001_Scope_of_registration_guidance_updated_March_2015_01.pdf)   Last accessed: 16th March 2019.

106.   Cambs   and   Peterborough   Clinical   Commissioning   Group. https://www.cambridgeshireandpeterboroughccg.nhs.uk/health-professionals/resources-and-guidance/infection-prevention-and-control Last accessed 28th February 2018.

107.   Public Health England, Manchester. Microbiology Services Handbook. 2016. (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/516422/Manchester_microbiology_services_handbook_110416.pdf)

108.   National Health Service. NHS five year forward view 2014. Available at: https://www.england.nhs.uk/wp-content/uploads/2014/10/5yfv-web.pdf. Last accessed: 16th March 2019.

109.   Bennett G, Mansell I. Universal precautions: a survey of community nurses' experience and practice. Journal of Clinical Nursing 2004;13:413-21.

110.   Nazarko L. Potential pitfalls in adherence to hand washing in the community. British Journal of Community Nursing 2009;14:64-8.

111.   Smith SM. A review of hand-washing techniques in primary care and community settings. Journal of Clinical Nursing 2009;18:786-90.

112.   Koser CU, Holden MT, Ellington MJ, et al. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. The New England Journal of Medicine 2012;366:2267-75.

113.   Kong Z, Zhao P, Liu H, et al. Whole-Genome Sequencing for the Investigation of a Hospital Outbreak of MRSA in China. PloS one 2016;11:e0149844.

114.   Harris SR, Cartwright EJ, Torok ME, et al. Whole-genome sequencing for analysis of an outbreak of meticillin-resistant *Staphylococcus aureus*: a descriptive study. The Lancet Infectious Diseases 2013;13:130-6.

115.   Azarian T, Cook RL, Johnson JA, et al. Whole-genome sequencing for outbreak investigations of methicillin-resistant *Staphylococcus aureus* in the neonatal intensive care unit: time for routine practice? Infection Control and Hospital Epidemiology 2015;36:777-85.

116.   Gordon NC, Pichon B, Golubchik T, et al. Whole-Genome Sequencing Reveals the Contribution of Long-Term Carriers in *Staphylococcus aureus* Outbreak Investigation. Journal of Clinical Microbiology 2017;55:2188-97.

117.   Sanger F, Air GM, Barrell BG, et al. The nucleotide sequence of bacteriophage phiX174. Nature 1977;265, 687-695.

118.   Sanger F. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. Journal of Molecular Biology 1975;94:441-8.

119.   Ledda A, Price JR, Cole K, et al. Re-emergence of methicillin susceptibility in a resistant lineage of *Staphylococcus aureus*. The Journal of Antimicrobial Chemotherapy 2017;72:1285-8.

120.   Liu L, Li Y, Li S, et al. Comparison of Next-Generation Sequencing Systems. Journal of Biomedicine and Biotechnology 2012:251364.

121.   Quick J, Ashton P, Calus S, et al. Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of Salmonella. Genome biology 2015;16:114.

122.   Martin J, Phan HTT, Findlay J, et al. Covert dissemination of carbapenemase-producing *Klebsiella pneumoniae* (KPC) in a successfully controlled outbreak: long- and short-read whole-genome sequencing demonstrate multiple genetic modes of transmission. The Journal of Antimicrobial Chemotherapy 2017;72:3025-34.

123.   Phan HTT, Stoesser N, Maciuca IE, et al. Illumina short-read and MinION long-read WGS to characterize the molecular epidemiology of an NDM-1 *Serratia marcescens* outbreak in Romania. The Journal of Antimicrobial Chemotherapy 2018;73(3):672-679.

124.   Harris SR, Feil EJ, Holden MT, et al. Evolution of MRSA during hospital transmission and intercontinental spread. Science 2010;327:469-74.

125.   McAdam PR, Templeton KE, Edwards GF, et al. Molecular tracing of the emergence, adaptation, and transmission of hospital-associated methicillin-resistant *Staphylococcus aureus*. Proceedings of the National Academy of Sciences of the United States of America 2012;109:9107-12.

126.   Holden MT, Hsu LY, Kurt K, et al. A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant *Staphylococcus aureus* pandemic. Genome Research 2013;23:653-64.

127.   Garvey MI, Pichon B, Bradley CW, et al. Improved understanding of an outbreak of meticillin-resistant *Staphylococcus aureus* in a regional burns centre via whole-genome sequencing. The Journal of Hospital Infection 2016;94:401-4.

128.   Ugolotti E, Di Marco E, Bandettini R, Biassoni R. Genomic characterization of a paediatric MRSA outbreak by next-generation sequencing. The Journal of Hospital Infection 2018;98:155-60.

129.   Torok ME, Harris SR, Cartwright EJ, et al. Zero tolerance for healthcare-associated MRSA bacteraemia: is it realistic? The Journal of Antimicrobial Chemotherapy 2014;69:2238-45.

130.   Department of Health and Social Care. Annual Report of the Chief Medical Officer, Volume Two, 2011. Infections and the rise of antimicrobial resistance. https://www.gov.uk/government/publications/chief-medical-officer-annual-report-volume-2 Last accessed 1st June 2019.

131. British Society for Antimicrobial Chemotherapy. The BSAC Bacteraemia Resistance Surveillance Programme Long-term Surveillance of the in vitro Activity of a Range of Antimicrobial Agents Against Potential Pathogens Isolated from Blood Samples of Patients with Clinically Significant Bacteraemia. 2018. http://www.bsacsurv.org/wp-content/uploads/2014/07/PROTOCOL-bact-2018_FINAL.pdf. Last accessed: 1st June 2019.

132. Grundmann H, Aanensen DM, van den Wijngaard CC, et al. Geographic distribution of *Staphylococcus aureus* causing invasive infections in Europe: a molecular-epidemiological analysis. PLoS medicine 2010;7:e1000215.

133. Aanensen DM, Feil EJ, Holden MT, et al. Whole-Genome Sequencing for Routine Pathogen Surveillance in Public Health: a Population Snapshot of Invasive *Staphylococcus aureus* in Europe. mBio 2016;7.

134. Anson LW, Chau K, Sanderson N, et al. DNA extraction from primary liquid blood cultures for bloodstream infection diagnosis using whole genome sequencing. Journal of Medical Microbiology 2018;67:347-57.

135. Argimon S, Abudahab K, Goater RJ, et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. Microbial genomics 2016;2:e000093.

136. Jolley KA, Maiden MC. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 2010;11:595.

137. Quan TP, Bawa Z, Foster D, et al. Evaluation of Whole-Genome Sequencing for Mycobacterial Species Identification and Drug Susceptibility Testing in a Clinical Setting: a Large-Scale Prospective Assessment of Performance against Line Probe Assays and Phenotyping. Journal of Clinical Microbiology 2018;56.

138. Gordon NC, Price JR, Cole K, et al. Prediction of *Staphylococcus aureus* antimicrobial resistance by whole-genome sequencing. Journal of Clinical Microbiology 2014;52:1182-91.

139. Bradley P, Gordon NC, Walker TM, et al. Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *Mycobacterium tuberculosis*. Nature Communications 2015;6:10063.

140. Fowler PW, Cole K, Gordon NC, et al. Robust Prediction of Resistance to Trimethoprim in *Staphylococcus aureus*. Cell Chemical Biology 2018.

141. Department of Health. MRSA Screening – Operational Guidance. 2010. http://webarchive.nationalarchives.gov.uk/20130123181012/http://www.dh.gov.uk/en/Publicationsandstatistics/Lettersandcirculars/Dearcolleagueletters/DH_114961.

142. Department of Health. Implementation of modified admission MRSA screening guidance for NHS, 2014. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/345144/Implementation_of_modified_admission_MRSA_screening_guidance_for_NHS.pdf. Last accessed: 1st June 2019.

143. Public Health England. UK Standards for Microbiology Investigations Identification of *Staphylococcus* species, *Micrococcus* species and *Rothia* species. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/375283/ID_7i3.pdf. Last accessed: 3rd May 2019.

144. Filippin L, Roisin S, Nonhoff C, Vandendriessche S, Heinrichs A, Denis O. Evaluation of the Automated Vitek 2 System for Detection of Various Mechanisms of Macrolide and Lincosamide Resistance in *Staphylococcus aureus*. Journal of Clinical Microbiology 2014;52:4087-9.

145. Coll F, Harrison EM, Toleman MS, et al. Longitudinal genomic surveillance of MRSA in the UK reveals transmission patterns in hospitals and the community. Science Translational Medicine 2017;9(413).

146.   Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome research 2008;18:821-9.

147.   Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 2006;22:2688-90.

148.   SS H, Yokoe DS, et al. Automated detection of infectious disease outbreaks in hospitals: a retrospective cohort study. PLoS Medicine 2010;23.

149.   Stachel A, Pinto G, Stelling J, et al. Implementation and evaluation of an automated surveillance system to detect hospital outbreak. American Journal of Infection Control 2017;45:1372-7.

150.   Russo PL, Shaban RZ, Macbeth D, et al. Impact of electronic healthcare-associated infection surveillance software on infection prevention resources: a systematic review of the literature. The Journal of Hospital Infection 2018;99:1-7.

151.   Kahl BC, Mellmann A, Deiwick S, et al. Variation of the polymorphic region X of the protein A gene during persistent airway infection of cystic fibrosis patients reflects two independent mechanisms of genetic change in *Staphylococcus aureus*. Journal of Clinical Microbiology 2005;43:502-5.

152.   Zhang K, McClure JA, Elsayed S, et al. Novel multiplex PCR assay for characterization and concomitant subtyping of staphylococcal cassette chromosome *mec* types I to V in methicillin-resistant *Staphylococcus aureus*. Journal of Clinical Microbiology 2005;43:5026-33.

153.   Boyle-Vavra S, Li X, Alam MT, et al. USA300 and USA500 Clonal Lineages of *Staphylococcus aureus* Do Not Produce a Capsular Polysaccharide Due to Conserved Mutations in the *cap5* Locus. mBio 2015;6.

154.   Reuter S, Ellington MJ, Cartwright EJ, et al. Rapid bacterial whole-genome sequencing to enhance diagnostic and public health microbiology. JAMA Internal Medicine 2013;173:1397-404.

155.   Diep BA, Gill SR, Chang RF, et al. Complete genome sequence of USA300, an epidemic clone of community-acquired meticillin-resistant *Staphylococcus aureus*. Lancet (London, England) 2006;367:731-9.

156.   Thurlow LR, Joshi GS, Richardson AR. Virulence strategies of the dominant USA300 lineage of community-associated methicillin-resistant *Staphylococcus aureus* (CA-MRSA). FEMS immunology and medical microbiology 2012;65:5-22.

157.   Planet PJ, LaRussa SJ, Dana A, et al. Emergence of the epidemic methicillin-resistant *Staphylococcus aureus* strain USA300 coincides with horizontal transfer of the arginine catabolic mobile element and *spe*G-mediated adaptations for survival on skin. mBio 2013;4:e00889-13.

158.   Thurlow LR, Joshi GS, Clark JR, et al. Functional modularity of the arginine catabolic mobile element contributes to the success of USA300 methicillin-resistant *Staphylococcus aureus*. Cell Host & Microbe 2013;13:100-7.

159.   Joshi GS, Spontak JS, Klapper DG, Richardson AR. Arginine catabolic mobile element encoded *speG* abrogates the unique hypersensitivity of *Staphylococcus aureus* to exogenous polyamines. Molecular Microbiology 2011;82:9-20.

160.   Gould FK, Brindle R, Chadwick PR, et al. Guidelines (2008) for the prophylaxis and treatment of methicillin-resistant *Staphylococcus aureus* (MRSA) infections in the United Kingdom. The Journal of Antimicrobial Chemotherapy 2009;63:849-61.

161.   Nathwani D, Morgan M, Masterton RG, et al. Guidelines for UK practice for the diagnosis and management of methicillin-resistant *Staphylococcus aureus* (MRSA) infections presenting in the community. The Journal of antimicrobial chemotherapy 2008;61:976-94.

162. Stevens DL, Bisno AL, Chambers HF, et al. Practice guidelines for the diagnosis and management of skin and soft tissue infections: 2014 update by the Infectious Diseases Society of America. Clinical Infectious Diseases 2014;59:e10-52.

163. Kearns AM, Ganner M, Hill RLR, et al. O118 Community-associated MRSA ST8-SCCmecIVa (USA-300): experience in England and Wales. International journal of antimicrobial agents;29:S27.

164. MRSA bacteraemia: monthly data by post infection review assignment. First published: 6th August 2014. https://www.gov.uk/government/statistics/mrsa-bacteraemia-monthly-data-by-post-infection-review-assignment. Last accessed: 7th March 2018.

165. Tosas Auguet O, Betley JR, Stabler RA, et al. Evidence for Community Transmission of Community-Associated but Not Health-Care-Associated Methicillin-resistant Staphylococcus Aureus Strains Linked to Social and Material Deprivation: Spatial Analysis of Cross-sectional Data. PLoS Medicine 2016;13:e1001944.

166. Dupieux C, Blonde R, Bouchiat C, et al. Community-acquired infections due to Staphylococcus argenteus lineage isolates harbouring the Panton-Valentine leucocidin, France, 2014. Euro surveillance 2015; 11;20(23).

167. Public Health England. Standards for Microbiology Investigations. Investigation of Specimens for Screening for MRSA. Issue date: 03.04.14. Last accessed: 7th October 2018.

168. Miller R, Walker AS, Knox K, et al. 'Feral' and 'wild'-type methicillin-resistant Staphylococcus aureus in the United Kingdom. Epidemiology and Infection 2010;138:655-65.

169. Mollaghan AM, Lucey B, Coffey A, Cotter L. Emergence of MRSA clone ST22 in healthy young adults in the community in the absence of risk factors. Epidemiology and Infection 2010;138:673-6.

170. Toleman MS, Reuter S, Coll F, Harrison EM, Peacock SJ. Local Persistence of Novel MRSA Lineage after Hospital Ward Outbreak, Cambridge, UK, 2011-2013. Emerging Infectious Diseases 2016;22:1658-9.

171. Planet PJ. Life After USA300: The Rise and Fall of a Superbug. The Journal of Infectious Diseases 2017;215:S71-S7.

172. Holmes A GM, McGuane S, Pitt TL, Cookson BD, Kearns AM. Staphylococcus aureus isolates carrying Panton-Valentine leucocidin genes in England and Wales: frequency, characterization, and association with clinical disease. Journal of Clinical Microbiology. 2005;43(5):2384-90.

173. Pichon B HR, Laurent F, Larsen AR, Skov RL, Holmes M, et al. Development of a real-time quadruplex PCR assay for simultaneous detection of nuc, Panton-Valentine leucocidin (PVL), mecA and homologue mecALGA251. The Journal of Antimicrobial Chemotherapy. 2012;67(10):2338-41.

174. Stamatakis A, Ludwig T, Meier H. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics 2005;21:456-63.

175. Toleman MS, Reuter S, Coll F, et al. Systematic Surveillance Detects Multiple Silent Introductions and Household Transmission of Methicillin-resistant Staphylococcus aureus USA300 in the East of England. The Journal of Infectious Diseases 2016;214:447-53.

176. Donker T, Reuter S, Scriberras J, et al. Population genetic structuring of methicillin-resistant Staphylococcus aureus clone EMRSA-15 within UK reflects patient referral patterns. Microbial Genomics 2017;3:e000113.

177. Seale AC, Gordon NC, Islam J, Peacock SJ, Scott JAG. AMR Surveillance in low and middle-income settings - A roadmap for participation in the Global Antimicrobial Surveillance System (GLASS). Wellcome Open Research 2017;2:92.

178. Packer S, Pichon B, Thompson S, et al. Clonal expansion of community-associated meticillin-resistant *Staphylococcus aureus* (MRSA) in people who inject drugs (PWID): prevalence, risk factors and molecular epidemiology, Bristol, United Kingdom, 2012 to 2017. Euro Surveillance 2019 Mar;24(13).

179. World Health Organization. Global Priority List of Antibiotic Resistant Bacteria to Guide Research, Discovery, and Development of New Antibiotics. 2017. http://www.who.int/medicines/publications/WHO-PPL-Short_Summary_25Feb-ET_NM_WHO.pdf. Last Accessed: 9th May 2018.

180. Pantosti A, Venditti M. What is MRSA? The European Respiratory Journal 2009;34:1190-6.

181. Sassi M, Felden B, Revest M, et al. An outbreak in intravenous drug users due to USA300 Latin-American variant community-acquired methicillin-resistant *Staphylococcus aureus* in France as early as 2007. European Journal of Clinical Microbiology & Infectious Diseases 2017;36:2495-501.

182. Bygott J, Enoch DA, Carson RP, et al. Presumed community-acquired meticillin-resistant *Staphylococcus aureus* (MRSA) isolates reflect spillover of healthcare-associated MRSA. The Journal of Hospital Infection 2008;69:197-8.

Chapter 9. Appendices

Chapter 2.

Appendix 2.1. "Cambridgeshire prospective study" by Coll *et al.:* Manuscript.

INFECTIOUS DISEASE

Longitudinal genomic surveillance of MRSA in the UK reveals transmission patterns in hospitals and the community

Francesc Coll,[1]* Ewan M. Harrison,[2] Michelle S. Toleman,[2,3,4] Sandra Reuter,[2] Kathy E. Raven,[2] Beth Blane,[2] Beverley Palmer,[5] A. Ruth M. Kappeler,[5,6] Nicholas M. Brown,[3,5] M. Estée Török,[2,3] Julian Parkhill,[4] Sharon J. Peacock[1,2,3,4]*

Genome sequencing has provided snapshots of the transmission of methicillin-resistant *Staphylococcus aureus* (MRSA) during suspected outbreaks in isolated hospital wards. Scale-up to populations is now required to establish the full potential of this technology for surveillance. We prospectively identified all individuals over a 12-month period who had at least one MRSA-positive sample processed by a routine diagnostic microbiology laboratory in the East of England, which received samples from three hospitals and 75 general practitioner (GP) practices. We sequenced at least 1 MRSA isolate from 1465 individuals (2282 MRSA isolates) and recorded epidemiological data. An integrated epidemiological and phylogenetic analysis revealed 173 transmission clusters containing between 2 and 44 cases and involving 598 people (40.8%). Of these, 118 clusters (371 people) involved hospital contacts alone, 27 clusters (72 people) involved community contacts alone, and 28 clusters (157 people) had both types of contact. Community- and hospital-associated MRSA lineages were equally capable of transmission in the community, with instances of spread in households, long-term care facilities, and GP practices. Our study provides a comprehensive picture of MRSA transmission in a sampled population of 1465 people and suggests the need to review existing infection control policy and practice.

INTRODUCTION

Staphylococcus aureus is responsible for a high proportion of community-associated invasive and soft tissue infections and is a leading cause of health care–associated infections (*1*). This burden is compounded by infection with methicillin-resistant *S. aureus* (MRSA), which results in increased mortality and hospitalization costs and longer hospital stays compared to methicillin-susceptible *S. aureus* infections (*2*). Successful reduction of MRSA infection rates depends on preventing MRSA transmission and detecting and containing outbreaks (*3*). Understanding the settings and circumstances under which MRSA evades current infection control measures is central to designing new strategies to reduce transmission.

MRSA carriage and infection have historically been associated with health care settings. Recent studies have demonstrated the value of applying whole-genome sequencing to define the spread of MRSA (*4–10*) and a range of other pathogens in hospitals. Whole-genome sequencing provides the ultimate resolution to discriminate between bacterial isolates and, when combined with epidemiological data, enables the reconstruction of transmission networks. Previous studies have largely focused on suspected outbreaks (*4–6*) or transmission in high-risk settings such as intensive care units (*7–10*). These snapshots have confirmed the potential of whole-genome sequencing to confirm or refute outbreaks, but the value that could be derived from applying this to entire populations, including those that bridge the divide between hospitals and the community, is unknown. Here, we report the findings of a 12-month prospective study of all MRSA-positive individuals detected by a large

diagnostic microbiology laboratory in the East of England in which an integrated analysis of epidemiological and sequence data provided a full picture of MRSA transmission.

RESULTS

Study participants and MRSA isolates

We identified 1465 MRSA-positive individuals in the East of England over a 12-month period (April 2012 to April 2013) by screening all samples submitted to a diagnostic microbiology laboratory by three hospitals and 75 general practitioner (GP) practices (see Fig. 1 for geographical distribution). Cases had a median age of 68 years [range, newborns to 101 years; interquartile range (IQR), 46 to 82 years]. We sequenced 2282 isolates cultured from their multisite screens (*n* = 1619) or diagnostic specimens (*n* = 663), which equated to 1 isolate from 1006 cases and a median of 2 isolates (range, 2 to 15; IQR, 2 to 3) from 459 cases (see Supplementary Materials and Methods for rationale for selecting isolates for sequencing and fig. S1 for number of isolates sequenced per case). About 80% of sequenced MRSA isolates were from samples submitted by the three study hospitals (1453 multisite screens and 372 diagnostic specimens), with the remainder submitted by GP practices (166 multisite screens and 291 diagnostic specimens). Multilocus sequence types (STs) were derived from sequence data, which revealed that most of the isolates belonged to clonal complex (CC) 22 (1667 of 2282, 73%), the predominant health care–associated lineage in the UK (*11*). This was followed in frequency by CC30 (*n* = 129, 5.6%), CC5 (*n* = 108, 4.7%), CC1 (*n* = 105, 4.6%), and CC8 (*n* = 87, 3.8%) (see table S1 for CC designation of the entire collection). Supplementary Materials and Methods provides a detailed description of the patient data collected, microbiology, sequencing methodology, and sequence data analyses, and fig. S2 shows a flowchart summarizing the data types used and analyses.

[1]London School of Hygiene and Tropical Medicine, London, UK. [2]University of Cambridge, Cambridge, UK. [3]Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK. [4]Wellcome Trust Sanger Institute, Cambridge, UK. [5]Public Health England, London, UK. [6]Papworth Hospital NHS Foundation Trust, Cambridge, UK.
*Corresponding author. Email: francesc.coll@lshtm.ac.uk (F.C.); sharon.peacock@lshtm.ac.uk (S.J.P.)

SCIENCE TRANSLATIONAL MEDICINE | REPORT
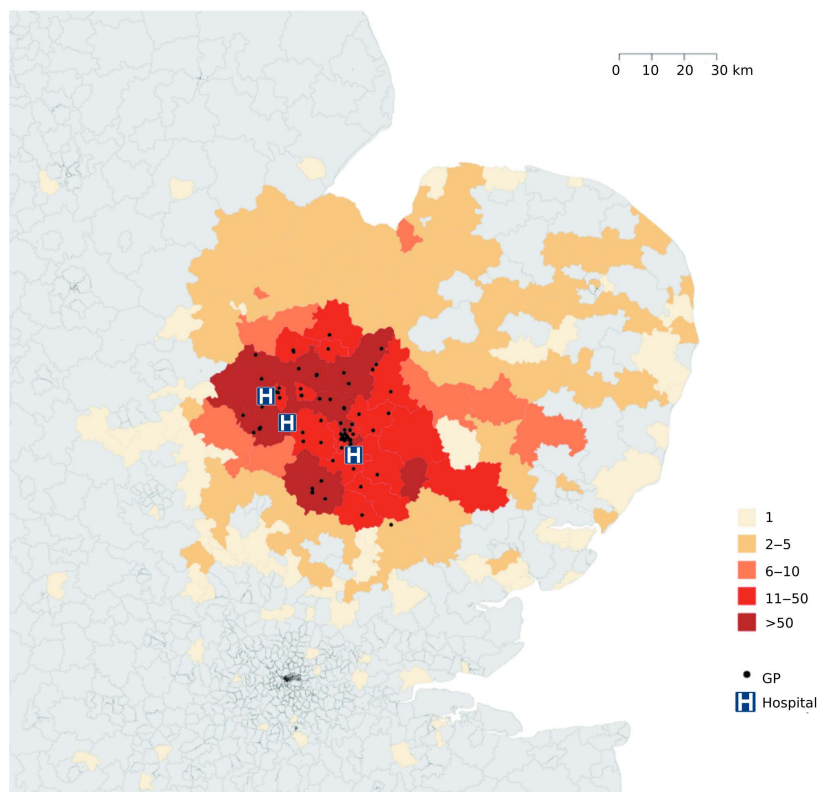
Fig. 1. Map showing the study catchment area in the East of England. The locations of hospitals (*n* = 3), GP practices (*n* = 75), and postcode districts are shown for the 1465 study cases. Postcode districts are color-coded to show the number of MRSA-positive cases sampled in each district. A total of 5,012,137 residents lived in the highlighted districts (16,240 km^2) according to the 2011 UK Census.

pital wards or other health care facilities with a high patient turnover or a proportionately higher prevalence of MRSA cases than the hospital- or community-averaged baseline. To assess the potential impact of this, we determined the strength of epidemiological links between people with genetically unrelated isolates (separated by more than 50 SNPs). This was achieved by a systematic pairwise comparison of 1040 cases with MRSA CC22. A total of 540,280 unique pairwise case comparisons were made, of which 534,417 had more than 50 SNPs (table S2). The instances of shared wards, GP practices, and postcodes were uncommon (wards/GP practices) or very rare (postcodes) for case pairs positive for unrelated CC22 MRSA (table S2). This analysis led us to classify shared postcodes (present in 0.04% of genetically unrelated cases), GP practice, and ward contacts (<1% of genetically unrelated cases) other than the Accident and Emergency Department (6.91%) as strong epidemiological links. Admission to the same hospital (particularly hospital A) was common in unrelated cases and considered a weak epidemiological link.

Each case was paired with the individual whose MRSA isolate was the closest genetic match, after which the genetic distance between each MRSA pair was plotted against six different categories of epidemiological contact (Fig. 2). This demonstrated a direct relationship between bacterial relatedness and strength of epidemiological contact.

Integration of genomic and epidemiological data

We initially divided the 2282 MRSA isolates into clusters containing isolates that were no more than 50 single-nucleotide polymorphisms (SNPs) different based on core genome comparisons (Supplementary Materials and Methods describes the rationale for the cutoff used). This led to the identification of 173 separate phylogenetic clusters. MRSA isolated from more than half of cases (785 of 1465, 53.6%) was genetically linked to MRSA from at least one other case based on isolates belonging to the same cluster. The next step was to apply epidemiological data (hospital admission and ward movement data, GP registration, and residential postcode) to this clustering framework to determine links between cases within each cluster, which ignored the traditional categorization of lineages as community- or hospital-associated. Figure S3 provides an overview of how the bacterial phylogeny and patient epidemiological data were integrated to define and classify transmission clusters. This revealed that 598 of 785 (76.2%) cases had an identifiable MRSA-positive contact with at least one other study case in a hospital setting and/or in the community (Table 1).

It is possible for epidemiological links between MRSA-positive individuals to arise by chance when MRSA carriers are admitted to hos-

Evidence of MRSA transmission in the community

Twelve percent of cases (72 of 598) with both bacterial and epidemiological links could be resolved into 27 distinct community transmission clusters. MRSA lineages regarded as community-associated (CA-MRSA)—which in the UK included CC1, CC5, CC8, CC45, and CC80—were associated with nine separate community transmission clusters (Table 1). However, most community clusters involved hospital-associated lineages [17 separate CC22 clusters involving 50 of 72 cases (69%) and 1 CC30 cluster involving 3 of 72 cases (4%)]. To contextualize the MRSA CC22 isolates associated with transmission in the community, we constructed a phylogenetic tree containing all CC22 study isolates. This showed that CC22 associated with community clusters was scattered throughout the phylogenetic tree, interspersed with clusters associated with cases with hospital contacts alone (Fig. 3). This indicates that CC22 isolates that were transmitted in the community belonged to the wider CC22 population, with no evidence for specific genetic subsets. We also identified transmission clusters relating to three independent GP practices, the largest of which contained 13 cases. All cases with shared postcodes were further investigated to determine whether they shared a residential address. This confirmed that MRSA transmission had occurred in at least 11 separate households (25 cases) and in

Table 1. Epidemiological classification of transmission clusters. Columns are ordered based on decreasing proportion of isolates in each CC. Each cell shows the number of cases and (in parentheses) the number of transmission clusters to which these cases were assigned. The number of transmission clusters in each category is the sum of those of its subcategories. The same applies to the number of cases except for columns "CC22" and "Overall." A total of seven cases had two different CC22 strains suggestive of mixed colonization or strain replacement that linked them to two different transmission clusters. This explains why the total number of genetically clustered cases ($n = 578$) is lower than the sum of cases in its subcategories. CCs with genetically unrelated isolates or identified in a single individual from the study population are not shown. "Multiple hospitals" refers to epidemiological contacts from more than one of the three study hospitals (A, B, and C).

Epidemiological classification	Overall	CC22	CC30	CC5	CC1	CC8	CC45	CC59	CC80	CC15	CC361
Genetically unrelated cases	680	462	36	49	35	42	17	15	6	1	2
Genetically clustered with other cases	785	578	46	30	45	9	34	26	9	8	3
Genetically clustered and epidemiological contacts	598 (173)	449 (127)	36 (8)	20 (9)	33 (13)	4 (2)	24 (8)	21 (3)	2 (1)	8 (1)	3 (1)
Only community contacts	72 (27)	50 (17)	3 (1)	3 (1)	6 (3)	4 (2)	4 (2)	—	2 (1)	—	—
Different postcode Shared GP practice	14 (3)	10 (1)	—	—	2 (1)	—	2 (1)	—	—	—	—
Same postcode Shared household	25 (11)	16 (7)	3 (1)	—	—	4 (2)	—	—	2 (1)	—	—
Same postcode Shared long-term care facility	22 (8)	20 (7)	—	—	—	—	2 (1)	—	—	—	—
Same postcode Different addresses	2 (1)	—	—	—	2 (1)	—	—	—	—	—	—
Same postcode Unresolved	9 (4)	4 (2)	—	3 (1)	2 (1)	—	—	—	—	—	—
Only hospital contacts	371 (118)	296 (91)	10 (3)	15 (7)	20 (8)	—	16 (5)	5 (2)	—	8 (1)	3 (1)
Ward contact	255 (64)	212 (52)	6 (1)	5 (2)	10 (4)	—	9 (2)	3 (1)	—	8 (1)	3 (1)
Hospital A	125 (41)	101 (35)	6 (1)	—	6 (2)	—	9 (2)	—	—	—	3 (1)
Hospital B	48 (14)	32 (10)	—	3 (1)	2 (1)	—	—	3 (1)	—	8 (1)	—
Hospital C	8 (4)	4 (2)	—	2 (1)	2 (1)	—	—	—	—	—	—
Multiple hospitals	75 (5)	75 (5)	—	—	—	—	—	—	—	—	—
Hospital-wide contact	118 (54)	85 (39)	4 (2)	10 (5)	10 (4)	—	7 (3)	2 (1)	—	—	—
Hospital A	97 (45)	70 (33)	2 (1)	8 (4)	8 (3)	—	7 (3)	2 (1)	—	—	—
Hospital B	6 (3)	2 (1)	2 (1)	—	2 (1)	—	—	—	—	—	—
Hospital C	8 (4)	6 (3)	—	2 (1)	—	—	—	—	—	—	—
Multiple hospitals	8 (2)	8 (2)	—	—	—	—	—	—	—	—	—
Both hospital and community contacts	156 (28)	104 (19)	23 (4)	2 (1)	7 (2)	—	4 (1)	16 (1)	—	—	—
Different postcode Shared GP practice	13 (2)	13 (2)	—	—	—	—	—	—	—	—	—
Same postcode Shared household	37 (9)	17 (3)	11 (3)	2 (1)	3 (1)	—	4 (1)	—	—	—	—
Same postcode Shared long-term care facility	56 (9)	36 (7)	—	—	4 (1)	—	—	16 (1)	—	—	—
Same postcode Different addresses	17 (3)	5 (2)	12 (1)	—	—	—	—	—	—	—	—
Same postcode Unresolved	33 (5)	33 (5)	—	—	—	—	—	—	—	—	—
Neither hospital nor community contacts	193	134	10	10	12	5	10	5	7	—	—
Total number of cases	1465	1040	82	79	80	51	51	41	15	9	5

8 long-term care facilities (22 cases) (Table 1). A pictorial representation of exemplars of transmission at a GP practice, long-term care facility, and household is shown in fig. S4 (A to C).

Evidence of MRSA transmission in hospitals

More than half of cases with epidemiological and bacterial genomic links (371 of 598, 62%) resided in transmission clusters with hospital contacts, of which 255 cases had ward contacts. The 371 cases were resolved into 118 different clusters each involving between 2 and 44 individuals (Table 1). We narrowed down further investigation to those clusters that contained five or more patients (nine clusters; see table S3 for details) and evaluated these for instances of direct ward contact (same ward, overlapping admission dates) or indirect ward contact (same ward, no overlap in admission dates). Where available,
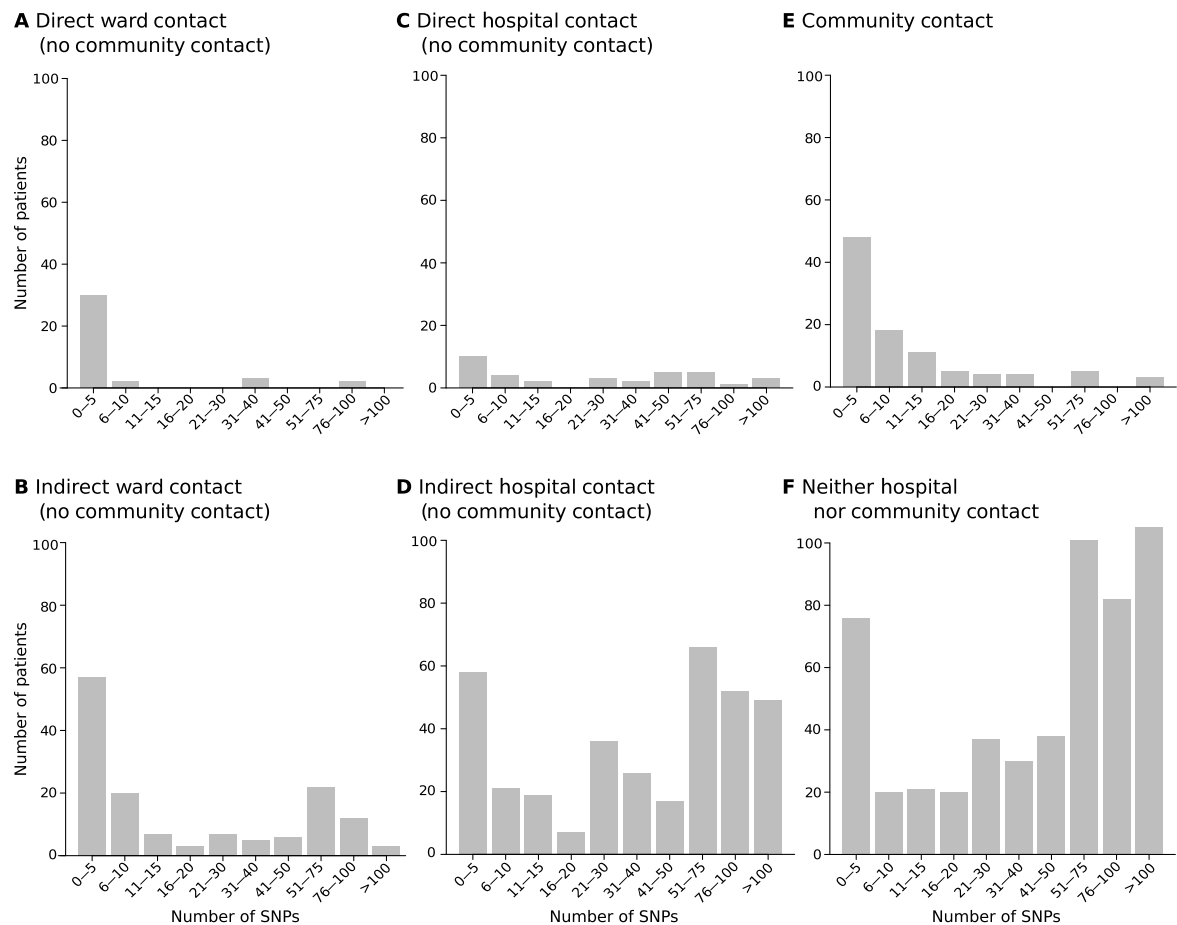
Fig. 2. Pairwise comparison between MRSA relatedness and type of patient contact. For each case, the most closely related MRSA isolate from another case was identified, and the epidemiological contact of each case pair was defined. The number of cases in each epidemiological category is shown as a function of the genetic distance (difference in the number of SNPs in the core genome). (**A** to **D**) Genetic distance distribution for cases with hospital contacts alone. Direct contact refers to a link in the same time and place (ward or hospital). Indirect contact refers to a link in the same place but different time. (**E**) Community contacts (shared residential postcodes or GP practice). (**F**) Cases with neither hospital nor community contacts. Only cases with MRSA isolates from CCs found in at least one other patient in the population are shown (n = 1459).

the presence of a negative MRSA culture followed by a positive MRSA culture was interpreted as additional evidence of hospital acquisition. The specific ward where MRSA had been putatively acquired could be determined in three of the nine clusters, one of which is depicted in Fig. 4A. This ward-centric pattern occurred in two different hospitals and across different CCs (CC22, CC30, and CC15). Notably, we observed that there was a time delay between presumptive acquisition date and first clinical detection of MRSA positivity in most cases (six of eight, three of four, and three of five patients). For the remaining six hospital clusters, multiple wards in the same hospital were plausible places of acquisition. We also observed a pattern of transmission that centered around specific individuals in which the movement of a single, persistently MRSA-positive index patient through multiple wards resulted in MRSA acquisition by numerous other patients. This

patient-centric pattern of transmission was identified in three transmission clusters (Fig. 4B and fig. S2, E and F) and was observed in two different hospitals and for two CCs (CC22 and CC30). Acquisition by other cases was associated with a high rate of indirect ward acquisition.

MRSA transmission at the hospital-community interface
We identified 28 clusters (157 cases) that contained a mixture of people with community and hospital epidemiological links (Table 1). Further analysis of 15 clusters that contained five or more cases (detailed in table S3) revealed instances of community-onset transmission followed by onward nosocomial dissemination, and hospital-onset transmission followed by nosocomial and community spread in CC30 and CC22 clusters. A pictorial representation of exemplars of these transmission patterns is shown in fig. S4 (D to F).

Coll *et al.*, *Sci. Transl. Med.* **9**, eaak9745 (2017)   25 October 2017

4 of 9

104 SNPs

■ Hospital transmission ■ Community transmission ■ Hospital/community transmission ■ No known epidemiological links
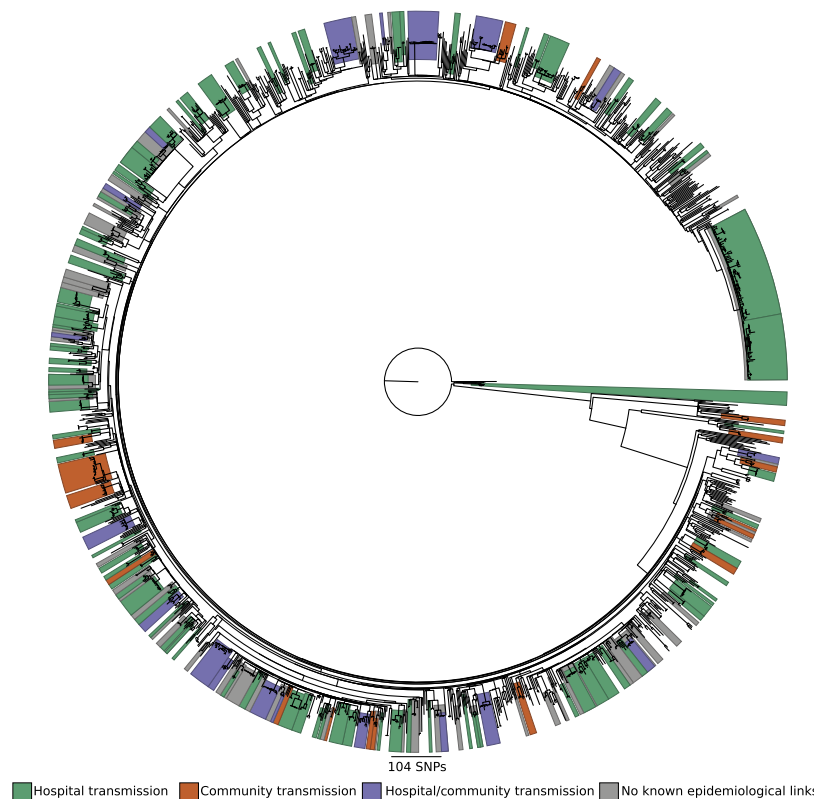
Fig. 3. Transmission clusters color-coded on the CC22 phylogeny. Maximum likelihood tree generated from 34,600 SNP sites in the core genome is shown for 1667 CC22 isolates. Colors refer to the type of epidemiological links in clusters of genetically related isolates (maximum 50 SNPs) from multiple cases.

DISCUSSION

Our findings have important implications for infection control policy and practice. MRSA transmission in our study population was not attributable to large nosocomial outbreaks but resulted from the cumulative effect of numerous clinically unrecognized episodes. We detected 173 separate genetic clusters that mapped to numerous different locations over the course of 12 months, which is indicative of repeated lapses in infection control. There are several explanations for extensive unrecognized transmission, including lack of hospital discharge swabbing and the fact that place of acquisition is often different to the place of detection and separated by a period of days, weeks, or months. This indicates the need for outbreak investigations to widen their scope in time and place when considering potential MRSA contacts.

Standard infection control practice centered on a ward-based approach may also fail to detect the impact of longitudinal patient-centric transmission. We identified a critical role for some persistent carriers who spread MRSA in multiple wards during complex health care pathways. This frequently involved indirect transmission, in which apparent acquisition by a new case occurred after the index case had left the ward, which is suggestive of environmental contamination or colonized health care workers. Further studies are needed to identify host factors responsible for persistent carriage associated with a high risk of MRSA trans-

mission to facilitate risk stratification and targeted allocation of isolation facilities where these are a limited resource.

It is generally accepted that most of the MRSA lineages either have become adapted to persist and spread in hospitals or are sufficiently fit to compete with other *S. aureus* lineages associated with community-associated carriage (*12*). CC22 is the predominant health care–associated MRSA lineage in the UK (~70%) followed in frequency by CC30, and most ongoing MRSA transmission is assumed to occur in health care settings. We expected that most clusters caused by CC22 and CC30 MRSA would map to hospitals but instead found considerable CC22 transmission in the community. Furthermore, clusters associated with community transmission of MRSA CC22 were distributed across the CC22 phylogeny and were interspersed with hospital-related clusters. This provides definitive evidence for the spread of so-called hospital-associated lineages such as CC22 through transmission networks that include the community. The repeated introduction of MRSA from the community into hospitals and vice versa signals the need for more robust action to detect and tackle community-associated carriage.

By including patient epidemiological information, we found that residential postcodes and GP registration information were strong epidemiological markers of MRSA transmission. Sharing the same postcode or GP practice by two or more MRSA-positive patients often

Coll *et al., Sci. Transl. Med.* **9**, eaak9745 (2017)   25 October 2017

5 of 9
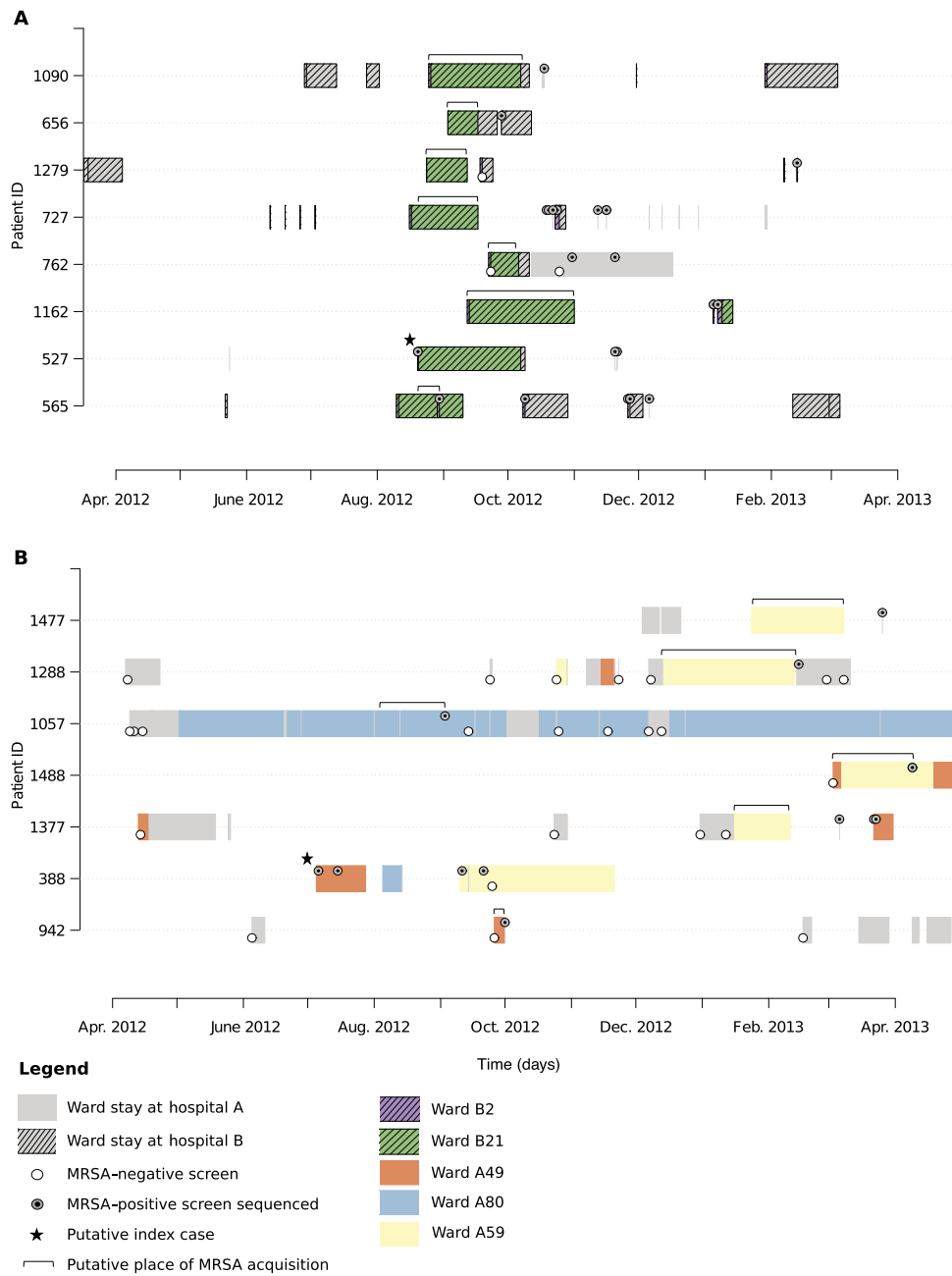
Fig. 4. Exemplars of two patterns of nosocomial MRSA spread. (**A**) Ward-centric pattern. Eight patients in this transmission cluster had ward contacts in wards B2 and B21, including admission overlaps. Notably, the putative epicenter of transmission was in ward B2 or B21, but the outbreak strain was isolated on later admissions in six of the eight patients, three of which (1090, 727, and 762) were first detected at a different hospital (hospital A) from where they had putatively acquired this strain (that is, in hospital B). (**B**) Patient-centric pattern. Six patients had stayed in wards visited by patient 388 (that is, A49, A80, and A59) before their MRSA isolation date. Negative MRSA screens before entry to these wards for some patients (1288, 1057, 1488, 1377, and 942) further support hospital acquisition. Isolates from patient 388 were the most basal in the phylogenetic tree, and their diversity enclosed that of isolates from the other patients, providing further indicators for this patient being the potential source for the transmission cluster. Colored blocks other than gray represent ward contacts, which are labeled by a letter to denote the hospital (A or B) and a number that denotes the anonymized ward.

Coll *et al.*, *Sci. Transl. Med.* **9**, eaak9745 (2017)   25 October 2017

SCIENCE TRANSLATIONAL MEDICINE | REPORT

indicated an outbreak, some of which spanned several months. Our findings support the routine collection of postcodes and GP registration as an integral part of routine surveillance to capture putative MRSA outbreaks in the community. This could guide a targeted approach to the use of whole-genome sequencing to confirm or refute transmission and direct infection control interventions that would curtail further dissemination.

We acknowledge several limitations of this study. The study design did not include longitudinal or discharge MRSA screening in hospitals or screening of environmental reservoirs and health care workers. Furthermore, sampling of the community was opportunistic and relied on samples submitted to the diagnostic microbiology laboratory. We acknowledge that this would mean failure to detect some MRSA carriers involved in our transmission clusters and that undetected carriers result in incomplete transmission routes being reconstructed. Nonsampled carriers explain why the MRSA isolate from 680 cases was not linked to the MRSA from any other case and why 193 cases whose isolate resided in a genetic cluster had no identifiable epidemiological contact. Despite detecting multiple transmission clusters, we are also likely to have underestimated the full extent of MRSA transmission attributable to nosocomial and community sources because of undersampling of the entire population served by the diagnostic laboratory at Cambridge University Hospitals.

In conclusion, we provide evidence for the value of integrated epidemiological and genomic surveillance of a population that accesses the same health care referral network in the East of England. The large number of patients screened here allowed us to sample MRSA lineages that are not dominant in the UK but are endemic in other areas of the world including USA300 (prevalent in the United States) (13), the European CA-MRSA CC80 (14), and the Taiwanese CC59 clone (prevalent in Asia) (15). The identification of transmission clusters involving these lineages in hospitals, in the community, and at the hospital-community interface suggests that our findings may be applicable to other UK regions and other countries.

MATERIALS AND METHODS

Study design

We conducted a 12-month prospective observational cohort study between April 2012 and April 2013 to identify consecutive individuals with MRSA-positive samples processed by the Clinical Microbiology and Public Health Laboratory at the Cambridge University Hospitals NHS Foundation Trust. This facility received samples from three hospitals (referred to as A, B, and C) and 75 GP practices in the East of England. All hospital inpatients were routinely screened for MRSA on admission to hospital, and screening was repeated weekly in critical care units. Compliance with mandatory admission screening at the three study hospitals was 85 to 90%. Additional clinical specimens were taken as part of routine clinical care. In the community, there was no formal MRSA screening, and specimens were taken by GPs or community nursing teams for clinical purposes, meaning that coverage was not complete. Epidemiological data (including hospital ward stays and residential postcodes) were recorded for all MRSA-positive cases. Detailed methodology is provided in Supplementary Materials and Methods, and a flowchart summarizing the data types and analyses undertaken is shown in fig. S2. The study protocol was approved by the National Research Ethics Service (reference 11/EE/0499), the National Information Governance Board Ethics and Confidentiality Committee (reference ECC 8-05(h)/2011), and the Cambridge University Hospitals

NHS Foundation Trust Research and Development Department (reference A092428).

DNA sequencing and genomic analyses

A total of 3053 MRSA isolates were collected during the study, of which 2320 were selected for whole-genome sequencing. A detailed description of the rationale for selecting isolates for sequencing and genomic methodologies is provided in Supplementary Materials and Methods. In brief, DNA was extracted, libraries were prepared, and 100–base pair paired end sequences were determined for 2320 isolates on an Illumina HiSeq2000, as previously described (11). Of these, 2282 were further analyzed after passing quality control (see Supplementary Materials and Methods). Genomes were de novo assembled using Velvet (16). STs were derived from assemblies, and CCs were assigned. All isolates assigned to the same CC were mapped using SMALT (www.sanger.ac.uk/science/tools/smalt-0) to the most closely related reference genome. SNPs were identified from BAM files using SAMtools (17). SNPs at regions annotated as mobile genetic elements were removed from whole-genome alignments, and maximum likelihood trees were created using RAxML (18) for each CC. Pairwise genetic distances between isolates of the same CC were calculated on the basis of the number of SNPs in the core genome. Sequence data were submitted to the European Nucleotide Archive (www.ebi.ac.uk/ena) under the accession numbers listed in data file S1.

Epidemiological analysis

We established epidemiological links between each pair of MRSA-positive individuals (termed case pairs) through a systematic comparison. Hospital contacts were categorized as follows: direct ward contact, if a case pair was admitted to the same ward with overlapping dates of admission; indirect ward contact, if admitted to the same ward with no overlapping dates; direct hospital-wide contact, if admitted to the same hospital in different wards with overlapping dates; and indirect hospital-wide contact, if admitted to the same hospital in different wards with no overlapping dates. We identified episodes of hospital admission for each case in the 12-month period before their first MRSA-positive sample. Information on outpatient clinic appointments was not available. Community contact was classified if cases shared a postcode or had their MRSA-positive sample submitted by the same GP practice. Community contacts were further categorized as follows: household contact, if people shared a residential address; long-term care facility contact, if they lived in the same long-term care facility; or GP contact, if they were registered with the same GP practice. Information on GP visits was not available other than that recorded for cases with MRSA swabs collected at GP practices. In a few instances, cases shared the same postcode but lived at a different residential address. In a minority of cases, patient addresses could not be retrieved from clinical records and were classified as "unresolved." We studied cases positive for MRSA CC22 to determine the frequency of different types of epidemiological contact among genetically unrelated cases, using a pairwise SNP distance greater than 50 SNPs. This analysis led us to consider epidemiological links as strong if they were ward contacts (other than Accident and Emergency visits), GP contacts, or shared postcodes, and weak if they were hospital-wide contacts and Accident and Emergency visits (see Supplementary Materials and Methods for details).

Identification of putative MRSA transmission

Selecting a SNP cutoff to define MRSA transmission clusters was informed by two independent lines of evidence. First, we established

the genetic diversity of the same MRSA clone in a single individual (pool of diversity) in 26 cases with more than one isolate (range, 2 to 3; median, 2) from independent samples cultured on the same day. The maximum genetic distance of MRSA in each case ranged from 0 to 41 SNPs (median, 2; IQR, 1 to 3), which is comparable to the maximum within-host diversity reported elsewhere (*19–21*). In parallel, we selected the single largest phylogenetic cluster containing isolates from cases with strong epidemiological links (13 cases, a putative outbreak) and established that the pairwise genetic distance between cases ranged from 0 to 48 SNPs. We constructed CC-based phylogenetic trees and then subdivided each tree into clusters based on a SNP distance of no more than 50 and looked for hospital and community contacts between cases residing in the same genetic cluster. Clusters were categorized as containing community contacts alone, hospital contacts alone, community and hospital contacts, or no known hospital/community contacts. For clusters with hospital and/or community contacts involving five or more cases, we incorporated individual patient movement data (for in-patients), sampling dates, MRSA screen results, and bacterial phylogeny to identify the most plausible MRSA source. Supplementary Materials and Methods and figs. S2 and S3 describe in more detail how genomic and epidemiological data were integrated to identify and classify transmission clusters.

SUPPLEMENTARY MATERIALS

www.sciencetranslationalmedicine.org/cgi/content/full/9/413/eaak9745/DC1
Materials and Methods
Fig. S1. Number of isolates sequenced per patient.
Fig. S2. Flowchart summarizing data types and analyses.
Fig. S3. Integration of genomic and epidemiological data to identify transmission clusters.
Fig. S4. Six examples of transmission clusters in different settings.
Fig. S5. Number of heterozygous sites in the core genome per isolate.
Fig. S6. Within-host diversity over time and at a single time point.
Table S1. Proportion of isolates in different CCs.
Table S2. Frequency of epidemiological contacts among genetically unrelated cases.
Table S3. Epidemiological classification of transmission clusters containing five or more cases.
Data file S1. Accession numbers.
References (*22–26*)

REFERENCES AND NOTES

1. F. D. Lowy, *Staphylococcus aureus* infections. *N. Engl. J. Med.* **339**, 520–532 (1998).
2. L. K. Yaw, J. O. Robinson, K. M. Ho, A comparison of long-term outcomes after meticillin-resistant and meticillin-sensitive *Staphylococcus aureus* bacteraemia: An observational cohort study. *Lancet Infect. Dis.* **14**, 967–975 (2014).
3. M. C. J. Bootsma, O. Diekmann, M. J. M. Bonten, Controlling methicillin-resistant *Staphylococcus aureus*: Quantifying the effects of interventions and rapid diagnostic testing. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5620–5625 (2006).
4. C. U. Köser, M. T. G. Holden, M. J. Ellington, E. J. P. Cartwright, N. M. Brown, A. L. Ogilvy-Stuart, L. Y. Hsu, C. Chewapreecha, N. J. Croucher, S. R. Harris, M. Sanders, M. C. Enright, G. Dougan, S. D. Bentley, J. Parkhill, L. J. Fraser, J. R. Betley, O. B. Schulz-Trieglaff, G. P. Smith, S. J. Peacock, Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *N. Engl. J. Med.* **366**, 2267–2275 (2012).
5. S. R. Harris, E. J. P. Cartwright, M. E. Török, M. T. G. Holden, N. M. Brown, A. L. Ogilvy-Stuart, M. J. Ellington, M. A. Quail, S. D. Bentley, J. Parkhill, S. J. Peacock, Whole-genome sequencing for analysis of an outbreak of meticillin-resistant *Staphylococcus aureus*: A descriptive study. *Lancet Infect. Dis.* **13**, 130–136 (2013).
6. L. Senn, O. Clerc, G. Zanetti, P. Basset, G. Prod'hom, N. C. Gordon, A. E. Sheppard, D. W. Crook, R. James, H. A. Thorpe, E. J. Feil, D. S. Blanc, The stealthy superbug: The role of asymptomatic enteric carriage in maintaining a long-term hospital outbreak of ST228 methicillin-resistant *Staphylococcus aureus*. *mBio* **7**, e02039-15 (2016).
7. U. Nübel, M. Nachtnebel, G. Falkenhorst, J. Benzler, J. Hecht, M. Kube, F. Bröcker, K. Moelling, C. Bührer, P. Gastmeier, B. Piening, M. Behnke, M. Dehnert, F. Layer, W. Witte, T. Eckmanns, MRSA transmission on a neonatal intensive care unit: Epidemiological and genome-based phylogenetic analyses. *PLOS ONE* **8**, e54898 (2013).
8. S. W. Long, S. B. Beres, R. J. Olsen, J. M. Musser, Absence of patient-to-patient intrahospital transmission of *Staphylococcus aureus* as determined by whole-genome sequencing. *mBio* **5**, e01692-14 (2014).
9. J. R. Price, T. Golubchik, K. Cole, D. J. Wilson, D. W. Crook, G. E. Thwaites, R. Bowden, A. S. Walker, T. E. A. Peto, J. Paul, M. J. Llewelyn, Whole-genome sequencing shows that patient-to-patient transmission rarely accounts for acquisition of *Staphylococcus aureus* in an intensive care unit. *Clin. Infect. Dis.* **58**, 609–618 (2014).
10. S. Y. C. Tong, M. T. G. Holden, E. K. Nickerson, B. S. Cooper, C. U. Köser, A. Cori, T. Jombart, S. Cauchemez, C. Fraser, V. Wuthiekanun, J. Thaipadungpanit, M. Hongsuwan, N. P. Day, D. Limmathurotsakul, J. Parkhill, S. J. Peacock, Genome sequencing defines phylogeny and spread of methicillin-resistant *Staphylococcus aureus* in a high transmission setting. *Genome Res.* **25**, 111–118 (2015).
11. S. Reuter, M. E. Török, M. T. G. Holden, R. Reynolds, K. E. Raven, B. Blane, T. Donker, S. D. Bentley, D. M. Aanensen, H. Grundmann, E. J. Feil, B. G. Spratt, J. Parkhill, S. J. Peacock, Building a genomic framework for prospective MRSA surveillance in the United Kingdom and the Republic of Ireland. *Genome Res.* **26**, 263–270 (2016).
12. J. Knox, A.-C. Uhlemann, F. D. Lowy, *Staphylococcus aureus* infections: Transmission within households and the community. *Trends Microbiol.* **23**, 437–444 (2015).
13. M. S. Toleman, S. Reuter, F. Coll, E. M. Harrison, B. Blane, N. M. Brown, M. E. Török, J. Parkhill, S. J. Peacock, Systematic surveillance detects multiple silent introductions and household transmission of methicillin-resistant *Staphylococcus aureus* USA300 in the East of England. *J. Infect. Dis.* **214**, 447–453 (2016).
14. M. Stegger, T. Wirth, P. S. Andersen, R. L. Skov, A. De Grassi, P. M. Simões, A. Tristan, A. Petersen, M. Aziz, K. Kiil, I. Cirković, E. E. Udo, R. del Campo, J. Vuopio-Varkila, N. Ahmad, S. Tokajian, G. Peters, F. Schaumburg, B. Olsson-Liljequist, M. Givskov, E. E. Driebe, H. E. Vigh, A. Shittu, N. Ramdani-Bougessa, J.-P. Rasigade, L. B. Price, F. Vandenesch, A. R. Larsen, F. Laurent, Origin and evolution of European community-acquired methicillin-resistant *Staphylococcus aureus*. *mBio* **5**, e01044-14 (2014).
15. M. J. Ward, M. Goncheva, E. Richardson, P. R. McAdam, E. Raftis, A. Kearns, R. S. Daum, M. Z. David, T. L. Lauderdale, G. F. Edwards, G. R. Nimmo, G. W. Coombs, X. Huijsdens, M. E. J. Woolhouse, J. R. Fitzgerald, Identification of source and sink populations for the emergence and global spread of the East-Asia clone of community-associated MRSA. *Genome Biol.* **17**, 160 (2016).
16. D. R. Zerbino, E. Birney, Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829 (2008).
17. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin; 1000 Genome Project Data Processing Subgroup, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
18. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
19. T. Golubchik, E. M. Batty, R. R. Miller, H. Farr, B. C. Young, H. Larner-Svensson, R. Fung, H. Godwin, K. Knox, A. Votintseva, R. G. Everitt, T. Street, M. Cule, C. L. C. Ip, X. Didelot, T. E. A. Peto, R. M. Harding, D. J. Wilson, D. W. Crook, R. Bowden, Within-host evolution of *Staphylococcus aureus* during asymptomatic carriage. *PLOS ONE* **8**, e61319 (2013).
20. O. C. Stine, S. Burrowes, S. David, J. K. Johnson, M.-C. Roghmann, Transmission clusters of methicillin-resistant *Staphylococcus aureus* in long-term care facilities based on whole-genome sequencing. *Infect. Control Hosp. Epidemiol.* **37**, 685–691 (2016).
21. G. K. Paterson, E. M. Harrison, G. G. R. Murray, J. J. Welch, J. H. Warland, M. T. G. Holden, F. J. E. Morgan, X. Ba, G. Koop, S. R. Harris, D. J. Maskell, S. J. Peacock, M. E. Herrtage, J. Parkhill, M. A. Holmes, Capturing the cloud of diversity reveals complexity and heterogeneity of MRSA carriage, infection and transmission. *Nat. Commun.* **6**, 6560 (2015).
22. M. Boetzer, C. V. Henkel, H. J. Jansen, D. Butler, W. Pirovano, Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578–579 (2011).
23. M. Boetzer, W. Pirovano, Toward almost closed genomes with GapFiller. *Genome Biol.* **13**, R56 (2012).
24. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
25. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
26. M. C. F. Prosperi, M. Ciccozzi, I. Fanti, F. Saladini, M. Pecorari, V. Borghi, S. Di Giambenedetto, B. Bruzzone, A. Capetti, A. Vivarelli, S. Rusconi, M. C. Re, M. R. Gismondo, L. Sighinolfi, R. R. Gray, M. Salemi, M. Zazzi, A. De Luca; ARCA collaborative group, A novel methodology for large-scale phylogeny partition. *Nat. Commun.* **2**, 321 (2011).

Coll *et al., Sci. Transl. Med.* **9**, eaak9745 (2017) 25 October 2017

8 of 9

141

SCIENCE TRANSLATIONAL MEDICINE | REPORT

Citation: F. Coll, E. M. Harrison, M. S. Toleman, S. Reuter, K. E. Raven, B. Blane, B. Palmer, A. R. M. Kappeler, N. M. Brown, M. E. Török, J. Parkhill, S. J. Peacock, Longitudinal genomic surveillance of MRSA in the UK reveals transmission patterns in hospitals and the community. *Sci. Transl. Med.* **9**, eaak9745 (2017).

Coll *et al., Sci. Transl. Med.* **9**, eaak9745 (2017)   25 October 2017   **9 of 9**

142

Chapter 3.

Appendix 3.1. Antibiotic susceptibility testing undertaken.

Antibiotic String Position Number	Antibiotic Agent
1	Benzylpenicillin
2	Cefoxitin
3	Oxacillin
4	Ciprofloxacin
5	Erythromycin
6	Chloramphenicol
7	Daptomycin
8	Fusidic acid
9	Gentamicin
10	Linezolid
11	Mupirocin
12	Nitrofurantoin
13	Rifampicin
14	Teicoplanin
15	Tetracycline
16	Tigecycline
17	Trimethoprim
18	Vancomycin
19	Clindamycin

Appendix 3.2. Codes for Antibiotic susceptibility profiles with >10 isolates determined in analysis.

Antibiotic Susceptibility Profile	Code	n	Frequency (%)
RRRRSSSSSSSSSSSSSSSS	ASP1	383	26.14
RRRRRSSSSSSSSSSSSSSS	ASP2	315	21.50
RRRSSSSRSSSSSSSSSSSS	ASP3	83	5.67
RRRSSSSSSSSSSSSSSSSS	ASP4	80	5.46
RRRRSSSRSSSSSSSSSSSS	ASP5	45	3.07
RRRSRSSSSSSSSSSSSSSS	ASP6	42	2.87
RRRRRSSRRSSSSSSSSSSS	ASP7	40	2.73
RRRRRSSRSSSSSSSSSSSS	ASP8	34	2.32
RRRSRSSRSSSSSSSSSSSS	ASP9	28	1.91
RRRRRSSSSSSSSSSSSRSS	ASP10	26	1.77
RRRRSSSSSSSSSSSSSRSS	ASP11	24	1.64
RRRRRSSSSSSSSSSRSSSS	ASP12	24	1.64
RRRRRSSSSSSSSSSSSSSR	ASP13	21	1.43
RRRRSSSRRSSSSSSSSSSS	ASP14	14	0.96
RRSSSSSSSSSSSSSSSSSS	ASP15	12	0.82
RRRSRSSSSSSSSSRSSSS	ASP16	12	0.82
RRRRRSSSSSSSSSSSISSSS	ASP17	12	0.82
RRRSSSSSSSSSSSSSRSS	ASP18	11	0.75
RRRRSSSSSSSSSSSISSSS	ASP19	11	0.75
RRRRRSSSSSSISSSSSSS	ASP20	10	0.68

Appendix 3.3. Phenotypic ASPs identified within *in silico spa*-types (For those *spa*-types with 10 or more isolates).

spa-type	Number of Isolates	Number of unique ASPs
Untypable	67	21
t002	40	16
t008	30	13
t018	19	9
t020	32	9
t022	55	11
t032	631	53
t127	56	14
t012	10	5
t019	33	5
t025	11	4
t1041	14	5
t1302	10	5
t316	35	8
t3505	10	4
t852	10	6
t3505	10	4
t379	17	5
t4545	36	5
t554	12	3

Chapter 4.

Appendix 4.1. Accession numbers and temporal data for isolates.
*Day 1: The first day of this study.

Study ID	Accession Number	Day Collected*
1	ERR204148	33
2	ERR737219	290
2	ERR715303	295
3	ERR736964	155
3	ERR714975	155
4	ERR701945	132
4	ERR730959	133
5	ERR737332	327
6	ERR213001	83
7	ERR715353	253
7	ERR737176	253
7	ERR715295	281
8	ERR714843	231
8	ERR702215	280
9	ERR212793	10
9	ERR702170	238
9	ERR702210	238
9	ERR702208	280
10	ERR736945	115
10	ERR730953	130
10	ERR774759	167
11	ERR737164	239
12	ERR737468	335
13	ERR736985	134
13	ERR702055	134
13	ERR715270	138
14	ERR737324	321
14	ERR715384	334
15	ERR737105	274
16	ERR737359	135
17	ERR715072	196
18	ERR715053	183
18	ERR737021	184
18	ERR736975	186
18	ERR736978	187
18	ERR715314	189
19	ERR737537	298

20	ERR715099	352
21	ERR714960	152
22	ERR212784	8
22	ERR742877	11
22	ERR737644	12
22	ERR715403	12
22	ERR715191	27
22	ERR715264	27
22	ERR715260	88
22	ERR715227	95
22	ERR715193	95
22	ERR715192	95
22	ERR715262	111
22	ERR742912	155
22	ERR715341	232
22	ERR702200	232
22	ERR737503	344
23	ERR702050	134
24	ERR701919	124

Chapter 5.

Appendix 5.1. Patient information letter.

DAY/MONTH/YEAR
Dear

We are writing to you as a patient registered at XXX who has received a positive result for meticillin resistant *Staphylococcus aureus* (MRSA) in the last 3-4 years.

The practice is currently working with the NHS, Public Health England (PHE) and the University of Cambridge to understand better how MRSA spreads in the community.

 As part of this, we will be looking in more detail at the medical records of a small number of our patients who are known to have had a positive result for a type of MRSA called ST22. Apart from basic data (such as age, sex, address), the key information that will be collected includes details of any contact you may have had with healthcare services and details of your MRSA.

The information collected will help us to determine whether our current systems are satisfactory or if anything more can be done to reduce the risk of people getting infections due to MRSA.

We would like your permission to share, with the review team, the relevant information from your medical records for the period of your MRSA infection and the 6 months before it was first detected. Any personal identifiable information will not be shared beyond the review team.

If you do not want us to share your information or would like further information, please get in touch with XXX by 11th August 2015. If we do not hear from you by this date, we will assume that you are happy for your information to be shared. Your decision will not influence your future care in any way.

Yours sincerely

Appendix 5.2. Case epidemiological data collection form.

Topic	Specific	
Patient information		
Identifiers	NHS number	
	Sample ID	
	First Name	
	Surname	
	Address	
	Postcode	
Demographics	Sex	
	DoB	
	Ethnicity	
MRSA information		
Sample information	**MRSA DETECTION DATE**	
	Sample location	
	Who took sample	
	Is this the first sample (during this episode) where MRSA was isolated from this patient	Yes or No
	If NO, please provide date of the first MRSA isolate and correct 6 month period below	
Characteristics of MRSA infection	Clinical information on MRSA infection	Free text
Outcome of MRSA infection		Free text
Patient health information		
Key conditions	List key health conditions	Free text
For the 6 month period PRE MRSA DETECTION DATE		
DATA COLLECTION PERIOD	**To be corrected to 6 month prior to first sample if necessary**	
Patient location in Pre Index period	House	Yes or No
	Nursing home/residential care	Yes or No
	Sheltered accommodation	Yes or No
	Other group living	Yes or No
	Address	
	Postcode	
	Dates moved if in period	
Skin integrity, did the patient have any breach in skin integrity	Pressure sore	Yes or No
	Leg ulcer	Yes or No

	Foot ulcer	Yes or No
	Eczema	Yes or No
	Surgical wound	Yes or No
	Chronic wound	Yes or No
	Other wound	Yes or No
	Other wound, describe	
	Other	Yes or No
	Other, describe	
Contact with, did the patient have contact with	**Any hospital inpatient stays?**	Yes or No
	Hospital	
	Date of admission	
	Date of Discharge	
	Dates	
	Reason	
	Any hospital outpatient appt?	Yes or No
	Hospital	
	Outpatient Clinic	
	Date	
	Reason	
	Any Hospital A&E appointment	Yes or No
	Hospital	
	Date	
	Reason	
	Dialysis	Yes or No
	Regularity	
	Date From	
	Date to	
	Community Podiatry	Yes or No
	Place	
	Date	
	GP Leg ulcer clinic	Yes or No
	Date	
	Other relevant GP clinics	Yes or No
	Name	
	Date	
	Any other GP visits	Yes or No
	GP	
	Date	
	Reason	
	Home visit by district nurse or other	Yes or No
	Nurse/team	
	Date	

	Reason	
	Continence Care	Yes or No
	Date	
	Other respite care	Yes or No
	Name	
	Dates	
Devices used	Type	List
	Dates	
	Duration	
	Purpose	
Presence of MRSA negative sample	Date of known negative MRSA Sample	e.g. previous clear screening sample
	Sample location	
	Who took sample	
Prior infections	Prior MRSA infection	
	Date of first sample	
	Date of decolonisation	
	Prior MSSA infection	
	Date of first sample	
	Date of decolonisation	
Connections	Known medical connection with another case e.g. same clinic	Free text
	Known family connection with another case	Free text
	Any other known connection with another case	Free text
Other information	Any other information relevant to the individual acquiring MRSA	Free text

Appendix 5.3. Staff screening information letter and consent form.

DAY/MONTH/YEAR

Dear colleague,

We are writing to you further to the meeting that was held with staff at XXXX on the 23rd of November. A research study by the University of Cambridge looked at the relatedness of routine MRSA samples in Cambridgeshire from April 2012 to April 2013 using whole-genome sequencing. This novel technology showed many groups of related MRSA samples, which would not have been identified previously. One of these groups included patients registered with XXXX.

Public Health England (PHE) and the University of Cambridge are working with the GP practice to use this information to advance our understanding of how MRSA spreads and how we can improve infection prevention and control beyond current practice. One approach is to undertake staff MRSA screening which involves taking swabs from the nose, throat (and optionally groin) and any areas of broken skin.

Further to the discussion at the practice meeting and on the advice of PHE, we are asking for staff who work on the XXXX premises to be screened. The screening will be undertaken confidentially. The clinical manager, XXXX, will organise for your swabs to be taken and samples will be forwarded to Addenbrooke's microbiology laboratory for processing. Where possible, swabs should be taken at the **start** of a shift to avoid detecting transient carriage of MRSA.

To preserve confidentiality, the letter with your results (positive or negative) will be sent to an address of your choice within two weeks of sampling. The results will not be sent directly to the XXXX or your GP. **The results will also be shared with the CCG Infection Prevention and Control (IPC) team to enable them to advise you further. If you are found to be MRSA positive you will be contacted by the CCG IPC team to provide you with further advice and details of decolonisation treatment.** Given the nature of work in primary care, it is highly unlikely that any work restrictions will be required.

We request that you provide a preferred contact telephone number for the team to use to contact you if you are identified as being MRSA positive (see form overleaf). It is also important to identify if an answer phone message can be left on a home or mobile telephone number if there is no answer, to avoid delay in contacting you.

Many thanks for your assistance with this screening. If you have any questions or queries about this process please feel free to contact me, or XXXX at the practice, for further advice.

Yours sincerely,

Dr Nicholas Brown, Consultant Microbiologist and Interim Lead Public Health Microbiologist, Public Health England, East of England.

Email: nicholas.brown@addenbrookes.nhs.uk
Telephone: 01223 257057

Staff MRSA Informed Consent, Screening Record and Contact Details

Informed consent
I agree to undergo MRSA screening and am aware that I will be contacted with the results.

Print name:

Signed: Date:

_____

Screening record

Individual taking swab

Print name:

Signed: Date of Swab:

_____

Contact Details

Name	
Date of Birth	
Job Title	
Date Screened	

Preferred address for results letter:

Preferred telephone number (if MRSA positive):

This is a
Home number /Mobile /Work number (please circle)

Can leave answer phone message if no answer Yes/ No (please circle)

_____
Please return this form to:
XXXXX

MRSA screening of staff: frequently asked questions

The information below is designed to answer staff questions about MRSA screening.

Why are we carrying out MRSA screening of staff? A research study by the University of Cambridge looked at the relatedness of routine MRSA samples in Cambridgeshire from April 2012 to April 2013 using whole-genome sequencing. This novel technology showed many groups of related MRSA samples, which would not have been identified previously. One of these groups included patients registered with XXXX. The GP surgery is working proactively with Public Health England (PHE) and University of Cambridge to understand how these findings from novel technologies can be best used to improve infection prevention and control beyond current practice.

Who will be screened? All staff using the premises of XXXX, including non-clinical staff, will be asked to have screening performed. There will also be environmental screening of the practice.

What does it involve? Screening involves a healthcare worker swabbing your nostrils, throat and groin at the beginning of a shift. The groin can be self-swabbed if you prefer. In addition, any open skin wounds should be swabbed.

How will I receive the results? When you undergo screening you will be asked to provide preferred contact details to receive the results. You will be contacted by a member of the team by phone if the result is positive to advise further. If the result is negative you will receive notification by letter within two weeks.

What does it mean if my MRSA test result is positive? If the test is positive MRSA decolonisation therapy will be required. MRSA treatment will only be required if you have signs of infection.

Will my employer be informed of this result? No, individual results are confidential between you and the PHE/CCG ICT team.

How will this affect my work? Given the nature of work in primary care it is highly unlikely that any work restrictions will be required. If decolonisation fails to eradicate the MRSA after several treatment courses, we may need to discuss the implications of this with you, occupational health and your employer. You will be informed in advance if this is necessary.

Will I pay for my treatments? No. Any treatment will be organised by the CCG Infection Prevention and Control team and you will not be charged for this.

How will this information be used? Summary data and feedback will be provided to the practice and staff. In addition summary data may be used in academic research settings to improve future practice. However no personal identifiable information or individual data will be disclosed.

Appendix 5.4. Environmental Sampling Sites.

SwabID	Location	Site
S01	Clinic Room 1	Lightswitch
S02	Clinic Room 1	Lamp handle
S03	Clinic Room 1	Skirting trim
S04	Clinic Room 1	Privacy Screen
S05	Clinic Room 1	Sink rim
S06	Clinic Room 1	Blind pulls
S07	Clinic Room 1	Lower couch
S08	Clinic Room 1	Footrest
S09	Clinic Room 1	Blood pressure (BP) cuff
S10	Clinic Room 1	Fan
S11	Clinic Room 1	Keyboard
S12	Clinic Room 1	Reusable scissors
S13	Clinic Room 1	Floor under couch
S14	Main waiting area	Carpet
S15	Main waiting area	Armrest of chair
S16	Main waiting area	Patient use BP monitor
S17	Main waiting area	Stored wheelchair frame
S18	Main waiting area	Check in monitor
S19	Main waiting area	Privacy screen
S20	Sub waiting area	Chair surface
S21	Clinic Room 2	Blind blade
S22	Clinic Room 2	Sink rim
S23	Clinic Room 2	Pillow
S24	Clinic Room 2	Specimen box
S25	Clinic Room 2	Notice board
S26	Clinic Room 2	Internal door handle
S27	Clinic Room 2	Soap dispenser surround
S28	Clinic Room 2	Horizontal shelving
S29	Clinic Room 3	Couch
S30	Clinic Room 3	Desktop
S31	Clinic Room 3	Horizontal shelving
S32	Clinic Room 3	BP cuff
S33	Clinic Room 3	Telephone
S34	Clinic Room 3	Communal KY jelly
S35	Treatment Room	Liquid Paraffin
S36	Treatment Room	Tongue depressor
S37	Clinic Room 4	Chair
S38	Clinic Room 4	Floor under couch
S39	Clinic Room 4	Glove box
S40	Clinic Room 4	Privacy curtain

Appendix 5.5. Accession numbers and assembly statistics for isolates.

StudyID	Accession Number	Total Length	No Contigs	Avg Contig Length	Largest Contig	N50	Contigs in N50
P01_1	ERS161495	2787852	28	99566.14	406415	310765	4
P02_1	ERS161347	2785773	31	89863.65	406345	176268	5
P02_2	ERS512442	2786952	26	107190.46	577341	235468	4
P03_1	ERS198012	2743760	28	97991.43	577983	174029	4
P03_2	ERS512836	2743767	23	119294.22	585540	210359	4
P04_1	ERS197550	2740977	26	105422.19	577179	210333	4
P04_2	ERS808624	2797195	26	107584.42	861382	275185	3
P04_3	ERS808620	2794537	27	103501.37	644342	325238	4
P04_4	ERS808623	2794805	25	111792.2	734450	337770	3
P05_1	ERS163223	2790656	24	116277.33	645061	210307	4
P06_1	ERS161486	2787317	31	89913.45	406600	176889	5
P06_2	ERS513104	2786014	22	126637	451170	210446	5
P07_1	ERS161496	2790396	29	96220.55	406413	235468	5
P07_2	ERS198039	2787148	27	103227.7	520222	252652	4
P07_3	ERS512702	2785763	26	107144.73	406772	235517	5
P08_1	ERS198000	2785894	27	103181.26	581078	235485	4
P09_1	ERS197548	2785559	25	111422.36	580944	235468	4
P10_1	ERS197832	2786945	30	92898.17	577273	310752	4
P11_1	ERS512443	2791175	31	90037.9	406499	174571	5
P11_2	ERS512963	2791562	27	103391.19	406422	310911	4
P12_1	ERS163210	2785842	29	96063.52	406739	210241	5
P12_2	ERS197734	2786143	33	84428.58	581210	235410	4
P13_1	ERS161475	2785024	28	99465.14	406601	173959	5
P13_2	ERS513007	2786489	26	107172.65	581322	235468	4
P13_3	ERS512501	2784896	26	107111.38	513426	252406	4
P14_1	ERS808627	2795284	23	121534.09	586873	253641	4
P15_1	ERS808626	2838387	30	94612.9	587033	235381	4
P16_1	ERS808619	2836067	29	97795.41	604554	384267	3
P17_1	ERS808621	2770001	23	120434.83	1029992	468119	2
P18_1	ERS808622	2747709	23	119465.61	508573	328798	4
P19_1	ERS808625	2774009	25	110960.36	1043699	659264	2
P20_1	ERS1066833	2755454	25	110218.16	1223185	1043875	2
P21_1	ERS1066834	2830409	39	72574.59	444283	178132	5
P22_1	ERS1066835	2814198	28	100507.07	628568	350510	3

Chapter 6.

Appendix 6.1. Accession numbers and assembly statistics for isolates.

Lane	Accession Number	Total Length	No Contigs	Avg Contig Length	Largest Contig	N50	Contigs in N50
12483_8_74	ERR527291	2805786	29	96751.24	578294	406874	3
12483_8_17	ERR527235	2785153	23	121093.61	710580	353665	3
12625_1_22	ERR541004	2789012	21	132810.1	680362	351019	3
12483_8_71	ERR527288	2827284	35	80779.54	336875	158444	7
12625_1_23	ERR541005	2786258	25	111450.32	410246	215664	5
12625_1_20	ERR541002	2892874	39	74176.26	310563	156340	7
12625_1_17	ERR540999	2898342	35	82809.77	331780	174046	7
12483_8_73	ERR527290	2783743	27	103101.59	430856	243617	5
12483_8_69	ERR527286	2904264	22	132012	712694	239538	4
12483_8_61	ERR527278	2800478	33	84862.97	406382	162908	6
12625_1_24	ERR541006	2791957	39	71588.64	328884	129459	7
12625_1_25	ERR541007	2954957	44	67158.11	355694	150949	7
12625_1_13	ERR540995	2748820	26	105723.85	578300	215897	4
12483_8_45	ERR527262	2814642	23	122375.74	442336	233965	4
12483_8_31	ERR527249	2849775	35	81422.14	334674	150786	6
12483_8_68	ERR527285	2808044	40	70201.1	307929	155404	7
12483_8_47	ERR527264	2951415	52	56757.98	252105	150360	8
12483_8_21	ERR527239	2782501	26	107019.27	526565	176875	5
12483_8_30	ERR527248	2827457	29	97498.52	654368	258448	4
12483_8_78	ERR527295	2825890	28	100924.64	340344	176999	6
12483_8_20	ERR527238	2827598	29	97503.38	499812	167026	6
12483_8_19	ERR527237	2743133	25	109725.32	493748	156373	5
12673_8_85	ERR555084	2822513	26	108558.19	360569	245050	5
12625_1_26	ERR541008	2946331	41	71861.73	354825	199458	6
12625_1_18	ERR541000	2805752	24	116906.33	536809	215896	4
12625_1_29	ERR541011	2810808	29	96924.41	540606	167877	5
12483_8_77	ERR527294	2831213	28	101114.75	406522	211865	5
12625_1_11	ERR540993	2777822	24	115742.58	517112	304025	4
12625_1_8	ERR540990	2806852	24	116952.17	857011	236909	3
12625_1_19	ERR541001	2827698	29	97506.83	426625	216095	5
12483_8_49	ERR527266	2845038	28	101608.5	637106	339932	3
12483_8_58	ERR527275	2795526	42	66560.14	363382	170598	6
12483_8_35	ERR527253	2841639	59	48163.37	395482	133886	7
12483_8_75	ERR527292	2826913	26	108727.42	449011	213417	5
12673_8_86	ERR555085	2834147	39	72670.44	548373	226039	4
12483_8_25	ERR527243	2813209	31	90748.68	406445	222865	5
12483_8_24	ERR527242	2840856	35	81167.31	289623	173973	6

12625_1_9	ERR540991	2808869	25	112354.76	597681	364517	3
12625_1_10	ERR540992	2811947	8	351493.38	1296603	1041310	2
12673_8_87	ERR555086	2915873	31	94060.42	410368	247491	5
12483_8_65	ERR527282	2799433	33	84831.3	439284	149389	5
12625_1_5	ERR540987	2796225	31	90200.81	339248	167285	6
12483_8_76	ERR527293	2830889	39	72586.9	340455	176721	6
12483_8_87	ERR527304	2891127	44	65707.43	310892	146370	7
12483_8_88	ERR527305	2792389	18	155132.72	1164948	381609	2
12483_8_93	ERR527310	2914920	30	97164	449946	343074	4
12483_8_2	ERR527220	2757398	25	110295.92	490637	253102	4
12483_8_52	ERR527269	2885056	36	80140.44	334661	174026	6
12483_8_83	ERR527300	2865262	42	68220.52	448536	211065	5
12673_8_88	ERR555087	2790152	29	96212.14	1043673	527546	2
12483_8_28	ERR527246	2854727	28	101954.54	406361	178507	6
12625_1_14	ERR540996	2832768	32	88524	348864	173980	6
12625_1_3	ERR540985	2825318	14	201808.43	548283	481656	3
12483_8_16	ERR527234	2799500	27	103685.19	538456	170552	6
12483_8_55	ERR527272	2803664	33	84959.52	404559	176944	5
12483_8_81	ERR527298	2936789	45	65261.98	253394	130847	8
12625_1_27	ERR541009	2833560	25	113342.4	572935	183521	5
12483_8_51	ERR527268	2861902	45	63597.82	371523	170377	6
12483_8_4	ERR527222	2798288	30	93276.27	312701	146422	7
12483_8_56	ERR527273	2866158	35	81890.23	334632	156330	6
12483_8_32	ERR527250	2827021	27	104704.48	532247	252836	4
12483_8_92	ERR527309	2826453	30	94215.1	335183	156386	6
12483_8_54	ERR527271	2787691	28	99560.39	400339	254900	5
12483_8_95	ERR527312	2857819	33	86600.58	318899	173966	6
12483_8_59	ERR527276	2844079	31	91744.48	406643	187194	5
12483_8_1	ERR527219	2800708	34	82373.76	530265	173960	5
12483_8_11	ERR527229	2823612	48	58825.25	404637	148370	6
12625_1_2	ERR540984	2857227	39	73262.23	312292	150786	7
12625_1_15	ERR540997	2805203	26	107892.42	531589	174151	4
12483_8_41	ERR527258	2802200	27	103785.19	539518	246114	4
12483_8_60	ERR527277	2794613	27	103504.19	579949	386206	3
12483_8_62	ERR527279	2800725	27	103730.56	406345	173968	5
12625_1_7	ERR540989	2870348	29	98977.52	335085	174567	6
12483_8_15	ERR527233	2808856	14	200632.57	1240701	544212	2
12483_8_7	ERR527225	2761492	29	95223.86	333533	176831	6
12625_1_4	ERR540986	2757384	27	102125.33	585437	407252	3
12483_8_80	ERR527297	2780864	16	173804	1012881	318029	3
12483_8_66	ERR527283	2782358	24	115931.58	625639	224810	4
12483_8_90	ERR527307	2731764	17	160692	826301	771413	2

12483_8_53	ERR527270	2829556	30	94318.53	417747	190560	5
12483_8_34	ERR527252	2823203	33	85551.61	410658	173968	5
12625_1_1	ERR540983	2831886	31	91351.16	406961	212595	5
12483_8_86	ERR527303	2945102	48	61356.29	355132	170904	6
12483_8_79	ERR527296	2746042	24	114418.42	662946	283591	4
12483_8_9	ERR527227	2841545	30	94718.17	432115	219258	5
12483_8_14	ERR527232	2784489	33	84378.45	360278	169827	5
12483_8_63	ERR527280	2745413	16	171588.31	957389	324483	3
12483_8_29	ERR527247	2904003	47	61787.3	253154	145784	8
12483_8_82	ERR527299	2859712	40	71492.8	406750	172821	6
12483_8_3	ERR527221	2852804	31	92025.94	339070	172516	6
12483_8_57	ERR527274	2784827	27	103141.74	445853	211703	5
12483_8_38	ERR527256	2787729	24	116155.38	443567	236587	4
12483_8_42	ERR527259	2769530	33	83925.15	528506	166712	6
12483_8_36	ERR527254	2726814	25	109072.56	333191	176658	6
12483_8_67	ERR527284	2782288	25	111291.52	406563	213079	5
12625_1_28	ERR541010	2874153	33	87095.55	443505	216245	5
12483_8_50	ERR527267	2803347	30	93444.9	406537	173991	5
12483_8_89	ERR527306	2741502	15	182766.8	607240	440589	3
12625_1_6	ERR540988	2795756	30	93191.87	625574	174882	5
12755_8_68	ERR564224	2821412	26	108515.85	511707	244583	5
12483_8_94	ERR527311	2884841	41	70361.98	334668	178065	6
12483_8_84	ERR527301	2803537	36	77876.03	348764	165123	6
12483_8_26	ERR527244	2768320	22	125832.73	1171791	349496	2
12625_1_30	ERR541012	2952602	41	72014.68	354364	170808	7
12625_1_32	ERR541014	2766459	16	172903.69	969638	766243	2
12589_1_6	ERR540707	2828482	24	117853.42	449190	211516	5
12625_1_33	ERR541015	2853499	33	86469.67	581872	210562	5
12625_1_34	ERR541016	2847832	19	149885.89	665585	222391	4
12625_1_35	ERR541017	2789449	39	71524.33	252887	122483	9
12625_1_36	ERR541018	2789751	28	99633.96	520827	176076	5
12625_1_37	ERR541019	2764775	27	102399.07	403134	187674	5
12625_1_38	ERR541020	2795876	15	186391.73	1259426	868969	2
12625_1_42	ERR541023	2816327	24	117346.96	641784	253084	4
12625_1_43	ERR541024	2817700	11	256154.55	1396697	607955	2
12625_1_45	ERR541026	2815609	30	93853.63	561845	236672	4
12625_1_46	ERR541027	2826754	29	97474.28	406603	215892	5
12625_1_47	ERR541028	2757611	26	106061.96	502337	176883	4
12625_1_48	ERR541029	2880371	31	92915.19	334637	176580	6
12625_1_49	ERR541030	2817986	38	74157.53	312907	169524	7
12625_1_53	ERR541034	2788288	27	103269.93	539951	406586	3
12625_1_54	ERR541035	2818505	24	117437.71	406733	168653	6

12625_1_50	ERR541031	2841089	33	86093.61	335264	190443	6
12625_1_51	ERR541032	2837131	23	123353.52	406750	211884	5
12625_1_52	ERR541033	2869983	40	71749.57	334427	174163	6
12625_1_55	ERR541036	2876246	19	151381.37	609282	421538	3
12625_1_56	ERR541037	2833519	33	85864.21	299986	159979	7
12625_1_57	ERR541038	2847960	26	109536.92	433732	336060	4
12625_1_59	ERR541040	2833232	28	101186.86	325018	190898	6
12625_1_60	ERR541041	2799996	22	127272.55	485656	236281	4
12625_1_61	ERR541042	2784951	27	103146.33	401456	211829	5
12625_1_62	ERR541043	2758550	21	131359.52	601319	254816	3
12625_1_63	ERR541044	2838019	26	109154.58	623662	190859	4
12625_1_64	ERR541045	2802106	29	96624.34	296088	183393	6
12625_1_65	ERR541046	2768537	15	184569.13	1004355	574665	2
12589_1_7	ERR540708	2863180	47	60918.72	326481	174569	6
12625_1_66	ERR541047	2829811	23	123035.26	521506	211546	5
12625_1_67	ERR541048	2929073	42	69739.83	323473	150304	8
12625_1_69	ERR541050	2889124	48	60190.08	307193	148843	7
12625_1_70	ERR541051	2795704	21	133128.76	532752	229302	5
12625_1_71	ERR541052	2812244	28	100437.29	495521	237888	4
12625_1_72	ERR541053	2728718	15	181914.53	1290868	485615	2
12625_1_73	ERR541054	2804579	28	100163.54	409160	183369	5
12625_1_74	ERR541055	2821148	31	91004.77	534430	185704	5
12625_1_75	ERR541056	2894237	19	152328.26	902815	616181	2
12625_1_76	ERR541057	2794704	24	116446	808154	398675	3
12625_1_78	ERR541059	2801895	14	200135.36	822244	644700	2
12625_1_79	ERR541060	2892772	31	93315.23	480959	170894	6
12625_1_80	ERR541061	2781800	18	154544.44	426918	305539	4
12625_1_81	ERR541062	2793711	14	199550.79	1043060	775249	2
12625_1_82	ERR541063	2932105	45	65157.89	354686	150369	8
12625_1_83	ERR541064	2765934	38	72787.74	300044	156345	7
12625_1_84	ERR541065	2808924	22	127678.36	426029	211888	5
12625_1_85	ERR541066	2809808	24	117075.33	406234	211846	5
12625_1_86	ERR541067	2774045	19	146002.37	622592	459067	3
12625_1_87	ERR541068	2792216	16	174513.5	1042516	398918	2
12625_1_89	ERR541070	2886035	34	84883.38	335172	146433	7
12625_1_90	ERR541071	2776797	48	57849.94	311566	156412	7
12625_1_91	ERR541072	2830918	27	104848.81	447811	325298	4
12625_1_92	ERR541073	2831198	28	101114.21	340338	174168	6
12625_1_93	ERR541074	2795747	30	93191.57	450586	179094	4
12625_1_95	ERR541076	2870876	30	95695.87	601992	252299	4
12593_1_1	ERR540795	2795381	17	164434.18	1028541	326924	3
12593_1_2	ERR540796	2801344	33	84889.21	354667	170582	6

12593_1 4	ERR540798	2818109	34	82885.56	406515	156349	6
12593_1 64	ERR540857	2729381	15	181958.73	764747	727015	2
12593_1 5	ERR540799	2802676	28	100095.57	406730	211567	5
12593_1 6	ERR540800	2946472	54	54564.3	252177	150846	8
12593_1 7	ERR540801	2843007	33	86151.73	334848	173924	7
12593_1 9	ERR540803	2802821	26	107800.81	407003	211279	5
12593_1 10	ERR540804	2817739	16	176108.69	1340092	326153	2
12593_1 11	ERR540805	2775205	35	79291.57	406736	156331	6
12593_1 13	ERR540807	2821445	29	97291.21	406652	311237	4
12593_1 65	ERR540858	2993887	43	69625.28	322311	169300	7
12593_1 15	ERR540809	2919161	21	139007.67	1229100	471733	2
12593_1 16	ERR540810	2722167	16	170135.44	1058188	598629	2
12593_1 17	ERR540811	2750792	35	78594.06	347780	224199	5
12593_1 18	ERR540812	2835115	32	88597.34	406288	213528	5
12593_1 19	ERR540813	2786940	33	84452.73	406171	160077	6
12593_1 20	ERR540814	2812598	13	216353.69	1393898	326231	2
12593_1 21	ERR540815	2791141	28	99683.61	422355	190485	5
12593_1 22	ERR540816	2854985	30	95166.17	406835	186271	5
12593_1 23	ERR540817	2883320	40	72083	310898	167257	7
12593_1 24	ERR540818	2802547	32	87579.59	339999	190453	6
12593_1 66	ERR540859	2927752	50	58555.04	326631	176815	6
12593_1 26	ERR540820	2875193	28	102685.46	335369	211429	6
12593_1 27	ERR540821	2874023	35	82114.94	337092	173974	6
12593_1 28	ERR540822	2881571	38	75830.82	338141	211446	5
12589_1 8	ERR540709	2834628	27	104986.22	406502	215472	5
12593_1 29	ERR540823	2749113	10	274911.3	1291662	544089	2
12593_1 30	ERR540824	2802886	38	73760.16	444231	211664	5
12593_1 68	ERR540861	2743349	34	80686.74	406910	174028	6
12593_1 69	ERR540862	2794465	28	99802.32	406743	211400	5
12589_1 9	ERR540710	2875276	43	66866.88	739265	387692	3
12589_1 10	ERR540711	2770288	17	162958.12	1047276	767560	2
12589_1 11	ERR540712	2818356	14	201311.14	1418352	1418352	1
12589_1 12	ERR540713	2861235	30	95374.5	495723	236603	4
12593_1 31	ERR540825	2802526	18	155695.89	1024385	386162	2
12589_1 13	ERR540714	2866277	36	79618.81	331774	156372	7
12593_1 32	ERR540826	2822711	15	188180.73	1396871	594472	2
12593_1 33	ERR540827	2829824	36	78606.22	390629	214664	5
12589_1 17	ERR540718	2808564	33	85108	582327	187028	5
12593_1 72	ERR540865	2820532	28	100733.29	540157	264183	4
12593_1 73	ERR540866	2817927	32	88060.22	406775	183667	5
12589_1 19	ERR540720	2830610	31	91310	449024	187597	5
12593_1 35	ERR540829	2852499	34	83897.03	354272	162431	6

12593_1_75	ERR540868	2758718	30	91957.27	314529	173974	6
12593_1_76	ERR540869	2845261	46	61853.5	356358	156399	7
12589_1_20	ERR540721	2797283	12	233106.92	981412	611731	2
12589_1_21	ERR540722	2798522	27	103648.96	414782	173970	5
12589_1_22	ERR540723	2829945	31	91288.55	334701	176674	6
12589_1_50	ERR540751	2869407	50	57388.14	406663	174046	6
12593_1_37	ERR540831	2832872	31	91382.97	306085	269181	5
12593_1_38	ERR540832	2803655	27	103839.07	406672	194148	5
12589_1_23	ERR540724	2785469	28	99481.04	406826	173959	6
12589_1_24	ERR540725	2791320	26	107358.46	383036	251837	5
12593_1_77	ERR540870	2836038	31	91485.1	580810	257862	4
12593_1_78	ERR540871	2906255	45	64583.44	275676	150854	7
12593_1_79	ERR540872	2907021	45	64600.47	385288	150867	7
12593_1_80	ERR540873	2799453	23	121715.35	428986	258032	4
12589_1_51	ERR540752	2823033	31	91065.58	368196	174562	6
12589_1_62	ERR540763	2858563	41	69721.05	272241	150354	7
12755_8_70	ERR564226	2806489	35	80185.4	411683	173986	5
12593_1_40	ERR540833	2952223	30	98407.43	745217	256581	4
12593_1_41	ERR540834	2773134	35	79232.4	349109	124766	7
12593_1_42	ERR540835	2849098	55	51801.78	328287	156449	7
12589_1_28	ERR540729	2734718	28	97668.5	525880	211643	4
12593_1_44	ERR540837	2808775	24	117032.29	409850	312987	4
12593_1_45	ERR540838	2803545	31	90436.94	355325	187753	5
12593_1_46	ERR540839	2900348	41	70740.2	252177	150367	8
12589_1_63	ERR540764	2744198	27	101636.96	356471	156386	6
12589_1_29	ERR540730	2791032	28	99679.71	581195	156333	5
12593_1_81	ERR540874	2826266	33	85644.42	327706	173972	6
12589_1_54	ERR540755	2918359	46	63442.59	300498	150835	7
12593_1_82	ERR540875	2795891	22	127085.95	495758	175610	5
12589_1_52	ERR540753	2758279	26	106087.65	731755	279248	3
12589_1_31	ERR540732	2790508	35	79728.8	325175	156338	6
12589_1_32	ERR540733	2851671	30	95055.7	335438	170729	6
12593_1_83	ERR540876	2739427	16	171214.19	642083	541567	3
12589_1_35	ERR540736	2846771	30	94892.37	425231	211425	5
12593_1_84	ERR540877	2772936	24	115539	680912	311055	3
12593_1_47	ERR540840	2810386	41	68546	406668	152545	6
12593_1_48	ERR540841	2785658	27	103172.52	375104	211785	5
12593_1_49	ERR540842	2813055	46	61153.37	254743	148162	8
12593_1_50	ERR540843	2815899	53	53130.17	365105	134203	7
12593_1_51	ERR540844	2790352	28	99655.43	605640	364398	3
12593_1_52	ERR540845	2787931	30	92931.03	348858	173985	6
12593_1_53	ERR540846	2803380	14	200241.43	988098	603253	2

12593_1_86	ERR540879	2795672	15	186378.13	1024696	437005	2
12593_1_87	ERR540880	2816771	16	176048.19	1339411	326299	2
12589_1_55	ERR540756	2834425	28	101229.46	409607	211681	5
12593_1_88	ERR540881	2791718	33	84597.52	333718	147001	6
12589_1_36	ERR540737	2859564	31	92244	448289	187790	5
12593_1_89	ERR540882	2799764	27	103694.96	297335	196766	6
12593_1_90	ERR540883	2887116	49	58920.73	298267	190165	6
12593_1_91	ERR540884	2753909	14	196707.79	1230710	1004294	2
12589_1_53	ERR540754	2784925	15	185661.67	776653	325728	3
12589_1_37	ERR540738	2841058	31	91647.03	409336	174028	5
12593_1_92	ERR540885	2766694	17	162746.71	1009743	718526	2
12589_1_38	ERR540739	2829756	36	78604.33	450656	187587	5
12589_1_39	ERR540740	2827983	36	78555.08	297474	146751	7
12593_1_93	ERR540886	2907459	32	90858.09	433624	203342	5
12589_1_40	ERR540741	2806788	28	100242.43	445379	214708	5
12593_1_95	ERR540888	2879278	36	79979.94	340240	131082	6
12589_1_1	ERR540702	2778244	28	99223	626119	375215	3
12589_1_2	ERR540703	2858688	51	56052.71	498406	126598	6
12589_1_42	ERR540743	2742514	17	161324.35	732743	327301	3
12589_1_43	ERR540744	2755754	21	131226.38	403326	236146	5
12589_1_44	ERR540745	2798775	30	93292.5	403038	215882	5
12589_1_3	ERR540704	2769245	11	251749.55	1295238	705716	2
12593_1_57	ERR540850	2754654	15	183643.6	601319	318708	4
12589_1_45	ERR540746	2841530	33	86106.97	296737	173971	6
12593_1_58	ERR540851	2814888	32	87965.25	534391	185703	4
12589_1_48	ERR540749	2943212	51	57710.04	252162	150923	8
12589_1_47	ERR540748	2787172	29	96109.38	406980	190408	5
12589_1_46	ERR540747	2796815	29	96441.9	402787	235908	5
12589_1_4	ERR540705	2834862	26	109033.15	623869	158607	4
12589_1_5	ERR540706	2827121	30	94237.37	531558	214877	4
12589_1_61	ERR540762	2786043	33	84425.55	317646	169740	6
12593_1_59	ERR540852	2747569	27	101761.81	579132	187334	4
12589_1_58	ERR540759	2847527	21	135596.52	858997	306322	3
12589_1_59	ERR540760	2747354	12	228946.17	1292715	1000816	2
12593_1_61	ERR540854	2850051	45	63334.47	296316	148376	7
12593_1_62	ERR540855	2766351	19	145597.42	931891	654638	2
12593_1_63	ERR540856	2919428	46	63465.83	300521	155132	7
12589_1_60	ERR540761	2737663	16	171103.94	775754	726968	2
12589_1_65	ERR540766	2840257	18	157792.06	1048269	561329	2
12589_1_66	ERR540767	2813103	17	165476.65	856984	811292	2
12589_1_67	ERR540768	2937977	54	54406.98	298545	142108	8
12589_1_68	ERR540769	2833910	32	88559.69	406835	215942	5

12589_1_69	ERR540770	2788662	30	92955.4	581087	158445	5
12589_1_70	ERR540771	2892904	35	82654.4	426153	173965	6
12589_1_71	ERR540772	2877622	36	79933.94	340067	177250	6
12589_1_72	ERR540773	2882132	44	65503	449078	174158	5
12589_1_73	ERR540774	2735694	14	195406.71	775721	723880	2
12589_1_74	ERR540775	2814049	14	201003.5	1339739	601644	2
12755_8_71	ERR564227	2865761	34	84287.09	319522	162406	6
12755_8_72	ERR564228	2818206	50	56364.12	253105	149712	7
12589_1_77	ERR540776	2828003	30	94266.77	354301	245290	5
12589_1_78	ERR540777	2805390	30	93513	562598	252894	4
12589_1_80	ERR540779	2864831	38	75390.29	388148	174028	6
12589_1_81	ERR540780	2798011	32	87437.84	554442	254705	4
12589_1_82	ERR540781	2810589	22	127754.05	406675	210561	5
12589_1_83	ERR540782	2859821	35	81709.17	424942	174158	5
12589_1_84	ERR540783	2866437	35	81898.2	407023	173965	6
12589_1_85	ERR540784	2860923	37	77322.24	433657	171480	5
12589_1_86	ERR540785	2818143	30	93938.1	578294	185011	5
12589_1_87	ERR540786	2865549	33	86834.82	331730	204066	6
12589_1_88	ERR540787	2755369	21	131208.05	866011	211058	3
12593_2_77	ERR540964	2817766	27	104361.7	532248	215969	4
12589_1_89	ERR540788	2858637	39	73298.38	310875	173809	6
12755_8_73	ERR564229	2845498	32	88921.81	406501	174013	5
12589_1_90	ERR540789	2874947	24	119789.46	506689	298007	4
12589_1_91	ERR540790	2842482	27	105277.11	310544	181666	6
12589_1_92	ERR540791	2864377	39	73445.56	335288	146496	6
12589_1_94	ERR540793	2810815	41	68556.46	311002	146329	7
12593_2_78	ERR540965	2757664	24	114902.67	578259	183012	4
12589_1_95	ERR540794	2746902	30	91563.4	586489	187386	4
12593_2_1	ERR540889	2784280	23	121055.65	435719	210501	5
12593_2_2	ERR540890	2884506	37	77959.62	336366	159531	6
12593_2_3	ERR540891	2869644	29	98953.24	342907	211791	5
12593_2_7	ERR540895	2736057	14	195432.64	674073	394939	3
12593_2_8	ERR540896	2822279	28	100795.68	923034	592632	2
12593_2_9	ERR540897	2783486	29	95982.28	403337	156332	5
12593_2_10	ERR540898	2796401	27	103570.41	432312	209690	5
12593_2_11	ERR540899	2734852	26	105186.62	601579	209205	4
12593_2_12	ERR540900	2850371	31	91947.45	447811	236605	4
12593_2_13	ERR540901	2873404	33	87072.85	308426	167280	7
12593_2_14	ERR540902	2875063	33	87123.12	391787	167183	6
12593_2_15	ERR540903	2796480	14	199748.57	1276109	393426	2
12593_2_16	ERR540904	2806264	24	116927.67	584728	190438	5
12593_2_17	ERR540905	2826252	47	60133.02	432174	125425	7

12593_2_18	ERR540906	2832323	34	83303.62	402665	173975	6
12593_2_20	ERR540908	2913436	41	71059.41	384115	224535	6
12593_2_21	ERR540909	2841399	33	86103	311221	155433	6
12593_2_22	ERR540910	2793063	15	186204.2	783665	327586	3
12593_2_25	ERR540913	2808505	31	90596.94	449202	252896	5
12593_2_79	ERR540966	2859996	34	84117.53	311794	156383	7
12593_2_26	ERR540914	2844388	35	81268.23	335264	180716	6
12593_2_27	ERR540915	2798768	31	90282.84	339954	145480	7
12593_2_80	ERR540967	2776565	15	185104.33	936644	405202	3
12593_2_28	ERR540916	2877882	43	66927.49	493057	222073	4
12593_2_74	ERR540961	2755223	26	105970.12	537138	270494	4
12593_2_29	ERR540917	2754003	24	114750.12	578231	198486	4
12593_2_30	ERR540918	2773695	33	84051.36	380518	172455	6
12593_2_31	ERR540919	2809265	26	108048.65	449758	198563	5
12593_2_34	ERR540922	2853024	38	75079.58	359249	156334	6
12593_2_40	ERR540927	2735446	16	170965.38	775746	721678	2
12593_2_42	ERR540929	2810521	33	85167.3	532086	253199	4
12593_2_43	ERR540930	2866408	33	86860.85	452294	198397	5
12593_2_37	ERR540925	2787135	34	81974.56	313355	156269	7
12593_2_38	ERR540926	2781323	23	120927.09	424252	288853	4
12593_2_45	ERR540932	2858091	24	119087.12	889981	347903	3
12593_2_46	ERR540933	2854137	39	73183	408720	174028	6
12593_2_47	ERR540934	2747192	25	109887.68	531601	429664	3
12593_2_48	ERR540935	2840984	44	64567.82	325722	173981	7
12593_2_49	ERR540936	2782682	25	111307.28	389949	180469	6
12593_2_50	ERR540937	2827626	29	97504.34	449052	161765	6
12593_2_51	ERR540938	2760287	9	306698.56	1170217	1018563	2
12593_2_52	ERR540939	2783598	30	92786.6	371722	209323	5
12593_2_53	ERR540940	2856170	38	75162.37	332380	146487	6
12593_2_54	ERR540941	2907721	52	55917.71	325147	172474	6
12593_2_55	ERR540942	2793753	23	121467.52	891425	227118	3
12593_2_56	ERR540943	2852512	35	81500.34	506997	170975	5
12593_2_81	ERR540968	2694936	24	112289	399842	169681	5
12593_2_61	ERR540948	2807408	33	85072.97	340311	174029	6
12593_2_62	ERR540949	2928689	32	91521.53	611252	311090	4
12593_2_82	ERR540969	2819094	27	104410.89	428301	185704	5
12593_2_64	ERR540951	2800241	17	164720.06	711581	355018	3
12593_2_57	ERR540944	2731988	12	227665.67	1091846	771520	2
12593_2_58	ERR540945	2797301	30	93243.37	571361	210567	4
12593_2_59	ERR540946	2828444	23	122975.83	1237435	426260	2
12593_2_60	ERR540947	2791233	39	71570.08	289393	172432	7
12593_2_69	ERR540956	2774566	31	89502.13	357087	238007	5

12593_2_70	ERR540957	2798601	30	93286.7	339941	176880	6
12593_2_71	ERR540958	2886555	37	78015	338798	209770	5
12593_2_83	ERR540970	2734638	18	151924.33	766419	726482	2
12593_2_72	ERR540959	2823402	33	85557.64	455215	167361	5
12593_2_65	ERR540952	2785432	27	103164.15	446139	212973	5
12593_2_84	ERR540971	2734798	13	210369.08	889306	721618	2
12593_2_66	ERR540953	2779769	31	89669.97	693654	331984	3
12593_2_67	ERR540954	2778851	13	213757.77	1287747	458670	2
12593_2_68	ERR540955	2950686	28	105381.64	714737	263489	4
12593_2_91	ERR540978	2902287	44	65961.07	252201	150384	8
12593_2_92	ERR540979	2784102	31	89809.74	404303	210634	5
12593_2_93	ERR540980	2762776	35	78936.46	581322	147926	6
12593_2_95	ERR540982	2861195	16	178824.69	1418386	542339	2
12673_8_44	ERR555043	2878255	18	159903.06	864716	366118	3
12673_8_78	ERR555077	2847565	36	79099.03	310572	158375	7
12673_8_60	ERR555059	2882444	22	131020.18	1010231	336319	3
12673_8_61	ERR555060	2786185	23	121138.48	670162	352485	3
12673_8_70	ERR555069	2952480	37	79796.76	714882	256733	4
12673_8_71	ERR555070	2876933	33	87179.79	450850	173973	5
12673_8_77	ERR555076	2898647	29	99953.34	611081	367375	4
12673_8_62	ERR555061	2915191	33	88339.12	464350	314988	4
12673_8_63	ERR555062	2772052	12	231004.33	1291631	543624	2
12673_8_66	ERR555065	2776146	17	163302.71	1090666	775789	2
12673_8_55	ERR555054	2769117	19	145743	387472	240923	5
12673_8_56	ERR555055	2827916	44	64270.82	338338	158318	7
12673_8_68	ERR555067	2882836	29	99408.14	505603	211932	4
12673_8_69	ERR555068	2798393	25	111935.72	409004	190333	5
12673_8_76	ERR555075	2813559	60	46892.65	277173	156412	7
12673_8_51	ERR555050	2816484	9	312942.67	1335735	544265	2
12673_8_52	ERR555051	2778989	35	79399.69	400703	174169	6
12673_8_79	ERR555078	2798664	42	66634.86	338698	125425	7
12673_8_80	ERR555079	2746006	14	196143.29	971061	584611	2
12673_8_81	ERR555080	2825296	25	113011.84	893943	255709	3
12673_8_54	ERR555053	2772753	13	213288.69	1291764	428796	2
12673_8_57	ERR555056	2833486	27	104943.93	406858	211675	5
12673_8_58	ERR555057	2737122	16	171070.12	766047	385595	3
12673_8_59	ERR555058	2710636	16	169414.75	1299005	625334	2
12593_2_85	ERR540972	2759201	28	98542.89	654828	174012	4
12673_8_45	ERR555044	2844311	35	81266.03	330651	166819	6
12673_8_46	ERR555045	2990078	66	45304.21	291353	111034	10
12673_8_48	ERR555047	2786913	26	107188.96	446091	245293	4
12673_8_47	ERR555046	2927433	34	86100.97	782091	298630	4

12673_8_50	ERR555049	2912386	46	63312.74	304400	150843	7
12673_8_53	ERR555052	2810797	54	52051.8	254996	120153	9
12673_8_72	ERR555071	2830164	43	65817.77	311105	146329	7
12673_8_73	ERR555072	2786692	31	89893.29	444390	209377	5
12673_8_74	ERR555073	2770516	36	76958.78	399519	134226	7
12593_2_87	ERR540974	2748159	26	105698.42	591719	174014	4
12593_2_88	ERR540975	2863268	36	79535.22	386184	190191	5
12593_2_89	ERR540976	2890755	42	68827.5	335252	255227	5
12593_2_90	ERR540977	2867185	39	73517.56	300290	170885	7
12593_2_86	ERR540973	2830337	36	78620.47	406473	168691	6

Appendix 6.2. Combined genomic and epidemiologic data.

Sample Details		Laboratory Data							
Strain ID	Year of Isolation	CC (genomic)	ST (genomic)	spa (genomic)	mecA	PVL	spa	Region	Community (CO)/ Hospital Onset(HO)
12483_8_74	2012	22	22	t608	1	0	t608	EAST	HO
12483_8_17	2012	5	5	t032	1	0	t032	S EAST	HO
12625_1_22	2012	5	5	t548	1	0	t548	YORK& HUM	CO
12483_8_71	2012	22	22	t032	1	0	t032	W MIDS	HO
12625_1_23	2012	22	3742	t906	1	0	t906	LONDON	HO
12625_1_20	2012	22	22	t032	1	0	t032	YORK& HUM	CO
12625_1_17	2012	22	22	t032	1	0	t032	LONDON	HO
12483_8_73	2012	5	125	t067	1	0	t067	LONDON	NK
12483_8_69	2012	8	8	t121	1	1	t121	EAST	CO
12483_8_61	2012	22	22	t032	1	0	t032	S EAST	CO
12625_1_24	2012	22	22	t005	1	1	t005	N WEST	HO
12625_1_25	2012	30	36	t1820	1	0	t1820	N WEST	HO
12625_1_13	2012	22	22	t022	1	0	t022	YORK& HUM	HO
12483_8_45	2012	22	22	t10125	1	0	t10125	N WEST	HO
12483_8_31	2012	22	22	t2892	1	0	t2892	S EAST	NK
12483_8_68	2012	22	22	t032	1	0	t032	S WEST	NK
12483_8_47	2012	30	36	t268	1	0	t268	NK	NK
12483_8_21	2012	22	22	t432	1	0	t432	LONDON	CO
12483_8_30	2012	22	22	t032	1	0	t032	N WEST	NK
12483_8_78	2012	22	22	t032	1	0	t032	S EAST	NK
12483_8_20	2012	22	22	t032	1	0	t032	S EAST	HO
12483_8_19	2012	22	22	t022	1	0	t022	LONDON	HO
12673_8_85	2012	22	22	t790	1	0	t790	LONDON	HO
12625_1_26	2012	30	36	t018	1	0	t018	S EAST	HO
12625_1_18	2012	22	22	t032	1	0	t032	LONDON	CO
12625_1_29	2012	22	22	t032	1	0	t032	N WEST	CO
12483_8_77	2012	22	22	t032	1	0	t032	N WEST	HO
12625_1_11	2012	59	59	t316	1	0	t316	W MIDS	HO
12625_1_8	2012	22	22	t379	1	0	t379	S EAST	HO
12625_1_19	2012	22	22	t032	1	0	t032	EAST	CO
12483_8_49	2012	22	22	t032	1	0	t032	EAST	CO
12483_8_58	2012	22	22	t032	1	0	t032	EAST	CO

12483_8_35	2012	22	22	t005	1	1	t005	LONDON	CO
12483_8_75	2012	22	22	t032	1	0	t032	S WEST	HO
12673_8_86	2012	8	94	t008	1	0	t008	LONDON	CO
12483_8_25	2012	22	22	t032	1	0	t032	N EAST	CO
12483_8_24	2012	22	22	t032	1	0	t032	S EAST	HO
12625_1_9	2012	22	22	t032	1	0	t032	S WEST	CO
12625_1_10	2012	12	12	t160	1	0	t160	S WEST	NK
12673_8_87	2012	8	8	t008	1	0	t008	LONDON	HO
12483_8_65	2012	59	59	t316	1	0	t316	YORK& HUM	HO
12625_1_5	2012	22	22	t1218	1	0	t1218	N WEST	CO
12483_8_76	2012	22	22	t032	1	0	t032	S EAST	HO
12483_8_87	2012	22	22	t032	1	0	t032	NK	NK
12483_8_88	2012	5	5	t032	1	0	t032	LONDON	CO
12483_8_93	2012	8	8	t008	1	0	t008	LONDON	HO
12483_8_2	2012	22	22	t785	1	0	t785	YORK& HUM	CO
12483_8_52	2012	22	22	t032	1	0	t032	NK	NK
12483_8_83	2012	22	22	t032	1	0	t032	E MIDS	HO
12673_8_88	2012	45	45	t032	1	0	t032	S WEST	HO
12483_8_28	2012	22	22	t906	1	0	t906	N WEST	HO
12625_1_14	2012	22	22	t032	1	0	t032	S EAST	HO
12625_1_3	2012	1	1	t127	1	0	t127	N WEST	CO
12483_8_16	2012	22	22	t032	1	0	t032	NK	NK
12483_8_55	2012	22	22	t032	1	0	t032	S EAST	HO
12483_8_81	2012	30	36	t018	1	0	t018	LONDON	HO
12625_1_27	2012	22	22	t032	1	0	t032	N EAST	HO
12483_8_51	2012	30	30	t1749	1	1	t1749	YORK& HUM	HO
12483_8_4	2012	22	22	t513	1	0	t513	S WEST	NK
12483_8_56	2012	22	22	- (Deletions/rearrangements within the spa gene. Loss of sequence complementary to forwar	1	0	t032	W MIDS	CO

				d primer)					
12483_8_32	2012	22	22	t2818	1	0	t2818	S WEST	CO
12483_8_92	2012	22	22	t578	1	0	t578	W MIDS	CO
12483_8_54	2012	22	22	t032	1	0	t032	S EAST	CO
12483_8_95	2012	22	22	t10172	1	0	t10172	S WEST	CO
12483_8_59	2012	22	22	t148	1	0	t148	W MIDS	CO
12483_8_1	2012	22	22	t032	1	0	t032	N WEST	CO
12483_8_11	2012	22	22	t022	1	0	t022	N WEST	HO
12625_1_2	2012	22	22	t032	1	0	t032	LONDON	HO
12625_1_15	2012	22	22	t032	1	0	t032	LONDON	HO
12483_8_41	2012	22	22	t1370	1	0	t1370	LONDON	HO
12483_8_60	2012	unknown CC	72	t032	1	0	t032	LONDON	HO
12483_8_62	2012	22	22	t032	1	0	t032	NK	NK
12625_1_7	2012	22	22	t032	1	0	t032	LONDON	CO
12483_8_15	2012	5	105	t045	1	0	t045	S WEST	HO
12483_8_7	2012	22	22	t022	1	0	t022	S WEST	NK
12625_1_4	2012	22	22	t4929	1	0	t4929	S WEST	NK
12483_8_80	2012	5	5	t6445	1	0	t6445	W MIDS	NK
12483_8_66	2012	30	30	t019	1	1	t019	LONDON	CO
12483_8_90	2012	5	5	t002	1	0	t002	S WEST	NK
12483_8_53	2012	22	22	t4218	1	0	t4218	S WEST	NK
12483_8_34	2012	22	22	t032	1	0	t032	LONDON	HO
12625_1_1	2012	22	22	t020	1	0	t020	W MIDS	HO
12483_8_86	2012	30	36	t018	1	0	t018	LONDON	HO
12483_8_79	2012	59	3669	t529	1	0	t529	S WEST	NK
12483_8_9	2012	22	22	t2857	1	0	t2857	W MIDS	NK
12483_8_14	2012	59	59	t316	1	0	t316	YORK& HUM	HO
12483_8_63	2012	5	5	t311	1	0	t311	EAST	CO
12483_8_29	2012	30	36	t018	1	0	t018	S WEST	HO
12483_8_82	2012	22	22	t022	1	0	t022	LONDON	HO
12483_8_3	2012	30	36	t253	1	0	t253	E MIDS	CO
12483_8_57	2012	22	22	t032	1	0	t032	E MIDS	CO
12483_8_38	2012	22	22	t032	1	0	t032	YORK& HUM	HO
12483_8_42	2012	22	22	t005	1	1	t005	LONDON	NK
12483_8_36	2012	22	22	t005	1	0	t005	YORK& HUM	HO

12483_8_67	2012	22	22	t032	1	0	t032	S EAST	CO
12625_1_28	2012	22	22	t032	1	0	t032	N WEST	HO
12483_8_50	2012	22	22	t022	1	0	t022	YORK& HUM	HO
12483_8_89	2012	5	5	t010	1	0	t010	E MIDS	CO
12625_1_6	2012	30	30	t019	1	1	t019	S EAST	CO
12755_8_68	2012	1	1	t127	1	1	t127	N WEST	CO
12483_8_94	2012	22	22	t032	1	0	t032	NK	NK
12483_8_84	2012	22	22	t020	1	0	t020	N WEST	NK
12483_8_26	2012	8	3727	t334	1	0	t334	S WEST	NK
12625_1_30	2012	30	36	t018	1	0	t018	LONDON	HO
12625_1_32	2012	5	3743	t1781	1	0	t1781	N WEST	CO
12589_1_6	2012	22	22	t3612	1	0	t3612	S WEST	CO
12625_1_33	2012	22	22	t020	1	0	t020	S WEST	CO
12625_1_34	2012	5	526	-(Deletions/rearrangements within the spa gene. Loss of sequence complementary to forward primer)	1	1	t002	N WEST	CO
12625_1_35	2012	22	22	t032	1	0	t032	EAST	HO
12625_1_36	2012	22	22	t032	1	0	t032	N EAST	NK
12625_1_37	2012	22	22	t032	1	0	t032	N EAST	NK
12625_1_38	2012	88	88	t1816	1	1	t1816	S EAST	NK
12625_1_42	2012	22	22	t8530	1	0	t8530	S WEST	CO
12625_1_43	2012	1	1	t127	1	0	t127	LONDON	HO
12625_1_45	2012	22	22	t11666	1	0	t11666	S WEST	CO
12625_1_46	2012	22	22	t3612	1	0	t3612	S WEST	CO
12625_1_47	2012	22	22	t022	1	0	t022	S WEST	CO
12625_1_48	2012	22	22	t432	1	0	t432	W MIDS	CO
12625_1_49	2012	22	22	t4573	1	1	t4573	W MIDS	CO
12625_1_53	2013	22	3122	t032	1	0	t032	S EAST	HO
12625_1_54	2013	22	22	t032	1	0	t032	S EAST	HO
12625_1_50	2012	22	22	t032	1	0	t032	N EAST	HO

12625_1_51	2013	22	22	t032	1	0	t032	W MIDS	CO
12625_1_52	2013	22	22	t032	1	0	t032	W MIDS	CO
12625_1_55	2013	8	8	t104	1	1	t104	S EAST	NK
12625_1_56	2012	22	22	t1041	1	0	t1041	EAST	CO
12625_1_57	2013	22	22	t8473	1	0	t8473	LONDON	HO
12625_1_59	2013	22	22	t032	1	0	t032	S EAST	HO
12625_1_60	2013	22	22	t032	1	0	t032	LONDON	HO
12625_1_61	2013	22	22	t032	1	0	t032	YORK& HUM	CO
12625_1_62	2013	22	22	t032	1	0	t032	LONDON	CO
12625_1_63	2012	22	22	t032	1	0	t032	N EAST	NK
12625_1_64	2013	22	22	t022	1	0	t022	LONDON	HO
12625_1_65	2013	unknown _CC	80	t044	1	1	t044	NK	NK
12589_1_7	2012	22	22	t032	1	0	t032	E MIDS	CO
12625_1_66	2013	22	22	t032	1	0	t032	E MIDS	HO
12625_1_67	2013	30	36	t018	1	0	t018	N WEST	HO
12625_1_69	2013	22	22	t032	1	0	t032	S EAST	CO
12625_1_70	2013	22	22	t906	1	0	t906	N WEST	CO
12625_1_71	2013	22	22	t1499	1	0	t1499	N WEST	NK
12625_1_72	2013	1	1	t127	1	0	t127	EAST	CO
12625_1_73	2013	22	22	t032	1	0	t032	S EAST	HO
12625_1_74	2013	22	22	t032	1	0	t032	LONDON	HO
12625_1_75	2013	8	8	t008	1	1	t008	YORK& HUM	CO
12625_1_76	2013	45	46	t040	1	0	t040	EAST	CO
12625_1_78	2013	unknown _CC	93	t202	1	1	t202	S EAST	CO
12625_1_79	2013	30	3674	t253	1	0	t253	E MIDS	HO
12625_1_80	2013	1	772	t657	1	1	t657	S WEST	NK
12625_1_81	2013	5	5	t002	1	1	t002	LONDON	HO
12625_1_82	2013	30	36	t018	1	0	t018	N WEST	CO
12625_1_83	2013	22	22	t1612	1	0	t1612	S WEST	NK
12625_1_84	2013	22	22	t032	1	0	t032	LONDON	CO
12625_1_85	2013	22	22	t032	1	0	t032	NK	NK
12625_1_86	2013	45	45	t077	1	0	t077	LONDON	CO
12625_1_87	2013	5	5	t002	1	1	t002	LONDON	HO
12625_1_89	2013	22	22	t032	1	0	t032	N WEST	CO
12625_1_90	2013	22	22	t852	1	1	t852	N WEST	HO
12625_1_91	2013	22	22	t032	1	0	t032	N WEST	CO
12625_1_92	2013	22	22	t032	1	0	t032	S EAST	HO

12625_1_93	2013	22	22	t020	1	0	t020	S EAST	CO
12625_1_95	2013	22	22	t032	1	0	t032	EAST	CO
12593_1_1	2013	5	5	t2724	1	0	t2724	N WEST	CO
12593_1_2	2013	22	22	t032	1	0	t032	S EAST	CO
12593_1_4	2013	22	22	t11885	1	0	t11885	LONDON	HO
12593_1_64	2013	5	5	t002	1	0	t002	LONDON	CO
12593_1_5	2013	22	22	t032	1	0	t032	E MIDS	CO
12593_1_6	2013	30	36	t018	1	0	t018	LONDON	HO
12593_1_7	2013	22	22	t11279	1	0	t11279	S EAST	NK
12593_1_9	2013	22	22	t020	1	0	t020	EAST	HO
12593_1_10	2013	1	1	t127	1	0	t127	N WEST	CO
12593_1_11	2013	22	22	t032	1	0	t032	N WEST	HO
12593_1_13	2013	22	22	t8964	1	0	t8964	EAST	CO
12593_1_65	2013	8	239	t037	1	0	t037	EAST	HO
12593_1_15	2013	8	8	t008	1	1	t008	LONDON	CO
12593_1_16	2013	5	5	t002	1	0	t002	N WEST	HO
12593_1_17	2013	59	59	t316	1	0	t316	W MIDS	CO
12593_1_18	2013	22	22	t022	1	0	t022	W MIDS	CO
12593_1_19	2013	22	22	t032	1	0	t032	N EAST	HO
12593_1_20	2013	1	1	t127	1	0	t127	LONDON	CO
12593_1_21	2013	22	22	t032	1	0	t032	S EAST	CO
12593_1_22	2013	22	22	t032	1	0	t032	LONDON	HO
12593_1_23	2013	22	22	t032	1	0	t032	LONDON	HO
12593_1_24	2013	22	22	t020	1	0	t020	EAST	NK
12593_1_66	2013	22	22	t020	1	0	t020	EAST	CO
12593_1_26	2013	22	22	t1467	1	0	t1467	N EAST	NK
12593_1_27	2013	22	22	t032	1	0	t032	EAST	HO
12593_1_28	2013	22	22	t3861	1	0	t3861	LONDON	CO
12589_1_8	2013	22	22	t9502	1	0	t9502	S WEST	HO
12593_1_29	2013	1	1	t2279	1	0	t2279	W MIDS	HO
12593_1_30	2013	22	22	t032	1	0	t032	S EAST	HO
12593_1_68	2013	22	22	t032	1	0	t032	LONDON	CO
12593_1_69	2013	22	22	t020	1	0	t020	LONDON	HO
12589_1_9	2013	5	5	t1341	1	0	t1341	S WEST	HO
12589_1_10	2013	5	1340	t002	1	0	t002	LONDON	CO
12589_1_11	2013	1	1	t127	1	0	t127	N WEST	HO
12589_1_12	2013	22	22	t022	1	0	t022	LONDON	HO
12593_1_31	2013	unknown CC	78	t186	1	0	t186	YORK& HUM	CO

12589_1_13	2013	22	22	t032	1	0	t032	LONDON	HO
12593_1_32	2013	5	5	t002	1	0	t002	LONDON	HO
12593_1_33	2013	22	22	t223	1	0	t223	W MIDS	CO
12589_1_17	2013	22	22	t032	1	0	t032	N EAST	NK
12593_1_72	2013	22	22	t2818	1	0	t2818	S WEST	NK
12593_1_73	2013	22	22	t5892	1	0	t5892	S EAST	HO
12589_1_19	2013	22	22	t032	1	0	t032	S WEST	NK
12593_1_35	2013	22	22	t032	1	0	t032	NK	NK
12593_1_75	2013	22	22	t032	1	0	t032	E MIDS	CO
12593_1_76	2013	22	22	t852	1	1	t852	LONDON	HO
12589_1_20	2013	5	5	t010	1	0	t010	EAST	CO
12589_1_21	2013	22	22	t032	1	0	t032	S EAST	HO
12589_1_22	2013	22	22	-(Deletions/rearrangements within the spa gene. Loss of sequence complementary to forward primer)	1	0	t032	W MIDS	CO
12589_1_50	2013	22	22	t032	1	0	t032	N WEST	HO
12593_1_37	2013	59	59	t316	1	0	t316	YORK& HUM	HO
12593_1_38	2013	22	22	t025	1	0	t025	N WEST	HO
12589_1_23	2013	22	22	t032	1	0	t032	LONDON	HO
12589_1_24	2013	unknown_CC	152	t1828	1	1	t1828	LONDON	CO
12593_1_77	2013	22	22	t022	1	0	t022	W MIDS	HO
12593_1_78	2013	30	36	t253	1	0	t253	S WEST	NK
12593_1_79	2013	30	36	t253	1	0	t253	S WEST	NK
12593_1_80	2013	22	2916	t578	1	0	t578	YORK& HUM	HO
12589_1_51	2013	22	22	t032	1	0	t032	LONDON	CO
12589_1_62	2013	30	36	t012	1	0	t012	LONDON	CO
12755_8_70	2013	22	22	t032	1	0	t032	E MIDS	CO
12593_1_40	2013	8	241	t037	1	0	t037	LONDON	CO
12593_1_41	2013	22	22	t4559	1	0	t4559	E MIDS	HO

12593_1_42	2013	22	22	t852	1	1	t852	LONDON	HO
12589_1_28	2013	22	22	t032	1	0	t032	S WEST	CO
12593_1_44	2013	22	22	t022	1	0	t022	N EAST	NK
12593_1_45	2013	22	22	t022	1	0	t022	N EAST	NK
12593_1_46	2013	30	2938	t018	1	0	t018	LONDON	CO
12589_1_63	2013	22	22	t1214	1	0	t1214	W MIDS	CO
12589_1_29	2013	22	3734	t12520	1	0	t12520	EAST	HO
12593_1_81	2013	22	3824	t022	1	0	t022	E MIDS	CO
12589_1_54	2013	30	3672	t018	1	0	t018	S EAST	HO
12593_1_82	2013	22	22	t032	1	0	t032	NK	NK
12589_1_52	2013	22	22	t032	1	0	t032	N WEST	HO
12589_1_31	2013	22	22	- (Deletions/rearrangements within the spa gene. Loss of sequence complementary to forward primer)	1	0	t032	E MIDS	CO
12589_1_32	2013	22	22	t020	1	0	t020	S EAST	HO
12593_1_83	2013	45	45	t630	1	0	t630	W MIDS	CO
12589_1_35	2013	22	22	t8473	1	0	t8473	LONDON	HO
12593_1_84	2013	22	22	t022	1	0	t022	S EAST	HO
12593_1_47	2013	22	22	t2818	1	0	t2818	S WEST	HO
12593_1_48	2013	22	22	t12254	1	0	t12254	W MIDS	NK
12593_1_49	2013	22	22	t557	1	0	t557	N EAST	HO
12593_1_50	2013	22	22	t852	1	1	t852	S WEST	CO
12593_1_51	2013	30	30	t019	1	1	t019	LONDON	CO
12593_1_52	2013	22	22	- (Deletions/rearrangements within the spa gene. Loss of sequence comple	1	0	t1218	N WEST	CO

				mentary to forward d primer)					
12593_1_53	2013	45	45	t026	1	0	t026	LONDON	CO
12593_1_86	2013	unknown_CC	78	t3202	1	0	t3202	YORK& HUM	HO
12593_1_87	2013	1	1	t127	1	0	t127	LONDON	HO
12589_1_55	2013	22	22	t032	1	0	t032	EAST	HO
12593_1_88	2013	22	22	t032	1	0	t032	YORK& HUM	NK
12589_1_36	2013	22	22	t6859	1	0	t6859	LONDON	NK
12593_1_89	2013	22	22	t022	1	0	t022	LONDON	CO
12593_1_90	2013	22	22	t032	1	0	t032	LONDON	CO
12593_1_91	2013	5	5	t002	1	0	t002	N WEST	HO
12589_1_53	2013	5	5	t002	1	1	t002	LONDON	HO
12589_1_37	2013	22	22	- (Deletions/rearrangements within the spa gene. Loss of sequence complementary to forward d primer)	1	0	t032	W MIDS	CO
12593_1_92	2013	5	5	t002	1	0	t002	N WEST	HO
12589_1_38	2013	22	22	t032	1	0	t032	N EAST	HO
12589_1_39	2013	22	3735	t12287	1	0	t12287	LONDON	HO
12593_1_93	2013	22	22	t12293	1	0	t12293	N EAST	NK
12589_1_40	2013	22	22	t379	1	0	t379	N EAST	NK
12593_1_95	2013	22	22	t032	1	0	t032	N WEST	HO
12589_1_1	2013	30	1456	t019	1	1	t019	LONDON	CO
12589_1_2	2013	45	45	t1081	1	0	t1081	LONDON	CO
12589_1_42	2013	5	3074	t002	1	0	t002	N WEST	CO
12589_1_43	2013	22	3734	t025	1	0	t025	EAST	HO
12589_1_44	2013	22	22	t557	1	0	t557	LONDON	HO
12589_1_3	2013	1	1	t127	1	0	t127	S EAST	CO
12593_1_57	2013	5	149	t002	1	0	t002	E MIDS	CO

12589_1_45	2013	22	22	t032	1	0	t032	W MIDS	HO
12593_1_58	2013	22	22	t032	1	0	t032	LONDON	CO
12589_1_48	2013	30	3671	t018	1	0	t018	LONDON	NK
12589_1_47	2013	22	22	t1214	1	0	t1214	LONDON	CO
12589_1_46	2013	22	22	t032	1	0	t032	LONDON	HO
12589_1_4	2013	22	22	t032	1	0	t032	NK	NK
12589_1_5	2013	22	22	t032	1	0	t032	W MIDS	CO
12589_1_61	2013	59	59	t7344	1	0	t7344	YORK& HUM	CO
12593_1_59	2013	22	22	t032	1	0	t032	YORK& HUM	HO
12589_1_58	2013	8	8	t1774	1	0	t1774	S WEST	NK
12589_1_59	2013	1	1	t127	1	0	t127	N WEST	HO
12593_1_61	2013	22	22	t032	1	0	t032	S EAST	CO
12593_1_62	2013	5	5	t002	1	0	t002	N WEST	HO
12593_1_63	2013	30	36	t018	1	0	t018	LONDON	HO
12589_1_60	2013	5	5	t002	1	0	t002	S WEST	CO
12589_1_65	2013	unknown_CC	3673	t044	1	1	t044	E MIDS	HO
12589_1_66	2013	5	5	t002	1	0	t002	S EAST	CO
12589_1_67	2013	30	36	t018	1	0	t018	S WEST	HO
12589_1_68	2013	22	22	t032	1	0	t032	S WEST	HO
12589_1_69	2013	22	22	t10718	1	0	t10718	N WEST	CO
12589_1_70	2013	22	22	t032	1	0	t032	LONDON	HO
12589_1_71	2013	22	22	t032	1	0	t032	S EAST	HO
12589_1_72	2013	22	22	t020	1	0	t020	S EAST	HO
12589_1_73	2013	5	5	t002	1	0	t002	NK	NK
12589_1_74	2013	1	1	t127	1	0	t127	LONDON	CO
12755_8_71	2013	22	22	t032	1	0	t032	N WEST	CO
12755_8_72	2013	22	22	t557	1	0	t557	N EAST	NK
12589_1_77	2013	22	22	t032	1	0	t032	N EAST	CO
12589_1_78	2013	22	22	t032	1	0	t032	N EAST	CO
12589_1_80	2013	22	22	t032	1	0	t032	NK	NK
12589_1_81	2013	22	22	t032	1	0	t032	LONDON	HO
12589_1_82	2013	22	22	t032	1	0	t032	LONDON	HO
12589_1_83	2013	22	22	t12422	1	0	t12422	LONDON	HO
12589_1_84	2013	22	22	t022	1	0	t022	LONDON	NK
12589_1_85	2012	22	22	t032	1	0	t032	LONDON	HO
12589_1_86	2013	22	3737	t032	1	0	t032	LONDON	HO
12589_1_87	2013	22	22	t032	1	0	t032	YORK& HUM	CO

12589_1_88	2013	22	22	t032	1	0	t032	N WEST	NK
12593_2_77	2013	22	22	t2235	1	0	t2235	S EAST	CO
12589_1_89	2013	22	22	t2752	1	0	t2752	YORK&HUM	CO
12755_8_73	2013	22	22	t032	1	0	t032	EAST	CO
12589_1_90	2013	8	8	t008	1	0	t008	LONDON	CO
12589_1_91	2013	22	1082	t1612	1	0	t1612	S EAST	CO
12589_1_92	2013	22	3738	t2818	1	0	t2818	S WEST	NK
12589_1_94	2013	22	22	t032	1	0	t032	W MIDS	HO
12593_2_78	2013	22	22	t020	1	0	t020	EAST	CO
12589_1_95	2013	22	22	t032	1	0	t032	YORK&HUM	CO
12593_2_1	2013	22	22	t022	1	0	t022	W MIDS	CO
12593_2_2	2013	22	22	t8703	1	0	t8703	LONDON	HO
12593_2_3	2013	22	22	t032	1	0	t032	S EAST	HO
12593_2_7	2013	97	97	t6576	1	0	t6576	S EAST	CO
12593_2_8	2013	5	5	t002	1	0	t002	LONDON	CO
12593_2_9	2013	22	22	t1302	1	0	t1302	N WEST	HO
12593_2_10	2013	22	22	t022	1	0	t022	LONDON	CO
12593_2_11	2013	59	59	t7344	1	0	t7344	YORK&HUM	CO
12593_2_12	2013	22	22	t032	1	0	t032	W MIDS	CO
12593_2_13	2013	22	22	t032	1	0	t032	N WEST	NK
12593_2_14	2013	22	22	t032	1	0	t032	N WEST	HO
12593_2_15	2013	88	88	t5973	1	0	t5973	N WEST	NK
12593_2_16	2013	22	22	t12550	1	0	t12550	N WEST	CO
12593_2_17	2013	22	22	t032	1	0	t032	E MIDS	HO
12593_2_18	2013	22	22	t022	1	0	t022	EAST	HO
12593_2_20	2013	30	36	t018	1	0	t018	S EAST	CO
12593_2_21	2013	22	22	t578	1	0	t578	N WEST	HO
12593_2_22	2013	5	526	-(Deletions/rearrangements within the spa gene. Loss of sequence complementary to forwar	1	0	t8084	S EAST	CO

				d primer)					
12593_2_25	2013	22	22	t032	1	0	t032	N EAST	NK
12593_2_79	2013	22	22	t032	1	0	t032	YORK& HUM	CO
12593_2_26	2013	22	22	t020	1	0	t020	S WEST	HO
12593_2_27	2013	22	3739	t032	1	0	t032	S WEST	HO
12593_2_80	2013	7	7	t12607	1	0	t12607	S WEST	CO
12593_2_28	2013	5	105	t002	1	0	t002	LONDON	CO
12593_2_74	2013	22	22	t032	1	0	t032	S EAST	HO
12593_2_29	2013	22	22	t032	1	0	t032	LONDON	HO
12593_2_30	2013	30	30	t021	1	1	t021	LONDON	CO
12593_2_31	2013	22	22	t12651	1	0	t12651	S WEST	HO
12593_2_34	2013	22	22	t020	1	0	t020	LONDON	CO
12593_2_40	2013	5	5	t002	1	0	t002	S WEST	CO
12593_2_42	2013	22	22	t022	1	0	t022	E MIDS	CO
12593_2_43	2013	22	22	t032	1	0	t032	E MIDS	HO
12593_2_37	2013	22	22	t1500	1	0	t1500	S WEST	HO
12593_2_38	2013	1	772	t657	1	1	t657	YORK& HUM	HO
12593_2_45	2013	8	8	t1774	1	0	t1774	N WEST	HO
12593_2_46	2013	22	22	t032	1	0	t032	S EAST	NK
12593_2_47	2013	22	22	t790	1	0	t790	EAST	HO
12593_2_48	2013	22	22	t032	1	0	t032	LONDON	CO
12593_2_49	2013	22	22	t032	1	0	t032	LONDON	NK
12593_2_50	2013	22	22	t025	1	0	t025	S WEST	NK
12593_2_51	2013	5	5	t1154	1	1	t1154	S EAST	CO
12593_2_52	2013	22	22	t032	1	0	t032	S WEST	HO
12593_2_53	2013	22	22	t032	1	0	t032	LONDON	HO
12593_2_54	2013	22	22	t032	1	0	t032	W MIDS	CO
12593_2_55	2013	30	30	t019	1	1	t019	S EAST	NK
12593_2_56	2013	30	36	t018	1	0	t018	YORK& HUM	HO
12593_2_81	2013	59	3740	t7344	1	0	t7344	YORK& HUM	CO
12593_2_61	2013	22	22	t032	1	0	t032	S EAST	HO
12593_2_62	2013	8	8	t008	1	0	t008	S EAST	HO
12593_2_82	2013	22	22	t020	1	0	t020	LONDON	HO
12593_2_64	2013	97	97	t12761	1	0	t12761	N WEST	HO
12593_2_57	2013	5	5	t002	1	0	t002	S WEST	NK

12593_2_58	2013	22	22	t022	1	0	t022	LONDON	CO
12593_2_59	2013	unknown_CC	361	t315	1	0	t315	LONDON	NK
12593_2_60	2013	22	22	t025	1	0	t025	S WEST	HO
12593_2_69	2013	22	22	t032	1	0	t032	LONDON	CO
12593_2_70	2013	22	22	t2892	1	0	t2892	S EAST	HO
12593_2_71	2013	22	22	t032	1	0	t032	W MIDS	CO
12593_2_83	2013	5	5	t002	1	0	t002	YORK& HUM	HO
12593_2_72	2013	22	22	t032	1	0	t032	LONDON	CO
12593_2_65	2013	22	22	t379	1	0	t379	W MIDS	HO
12593_2_84	2013	5	5	t002	1	0	t002	S WEST	CO
12593_2_66	2013	30	30	t122	1	1	t122	LONDON	NK
12593_2_67	2013	1	1	t127	1	0	t127	W MIDS	CO
12593_2_68	2013	8	241	t037	1	0	t037	LONDON	HO
12593_2_91	2013	30	36	t018	1	0	t018	LONDON	NK
12593_2_92	2013	22	22	t032	1	0	t032	LONDON	HO
12593_2_93	2013	22	22	t032	1	0	t032	E MIDS	CO
12593_2_95	2013	1	1	t127	1	0	t127	S EAST	HO
12673_8_44	2013	8	8	t008	1	1	t008	W MIDS	NK
12673_8_78	2013	22	3757	t032	1	0	t032	LONDON	CO
12673_8_60	2013	8	8	t008	1	1	t008	S EAST	HO
12673_8_61	2013	unknown_CC	361	t315	1	0	t315	LONDON	CO
12673_8_70	2013	8	241	t037	1	0	t037	LONDON	NK
12673_8_71	2013	22	22	t020	1	0	t020	S WEST	HO
12673_8_77	2013	8	8	t008	1	0	t008	S EAST	CO
12673_8_62	2013	8	8	t008	1	0	t008	YORK& HUM	CO
12673_8_63	2013	1	1	t127	1	0	t127	S WEST	CO
12673_8_66	2013	5	5	t002	1	0	t002	S WEST	NK
12673_8_55	2013	5	149	t002	1	0	t002	LONDON	NK
12673_8_56	2013	22	22	t032	1	0	t032	S EAST	NK
12673_8_68	2013	5	5	t002	1	0	t002	E MIDS	HO
12673_8_69	2013	22	22	t032	1	0	t032	E MIDS	HO
12673_8_76	2013	22	22	t005	1	1	t005	S EAST	CO
12673_8_51	2013	1	1	t127	1	0	t127	S WEST	HO
12673_8_52	2013	22	22	t032	1	0	t032	YORK& HUM	CO
12673_8_79	2013	22	22	t032	1	0	t032	N WEST	NK
12673_8_80	2013	88	88	t729	1	0	t729	S EAST	CO
12673_8_81	2013	5	5	t002	1	0	t002	S EAST	NK

12673_8_54	2013	1	1	t127	1	0	t127	S WEST	HO
12673_8_57	2013	22	22	t032	1	0	t032	E MIDS	HO
12673_8_58	2013	5	5	t002	1	0	t002	EAST	HO
12673_8_59	2013	45	3846	t015	1	0	t015	W MIDS	HO
12593_2_85	2013	22	22	t022	1	0	t022	N EAST	NK
12673_8_45	2013	22	22	t032	1	0	t032	E MIDS	HO
12673_8_46	2013	8	241	t030	1	0	t030	N WEST	HO
12673_8_48	2013	22	22	t032	1	0	t032	N WEST	NK
12673_8_47	2013	8	8	t032	1	0	t032	S EAST	HO
12673_8_50	2013	30	36	t018	1	0	t018	S EAST	CO
12673_8_53	2013	22	22	t005	1	1	t005	S EAST	HO
12673_8_72	2013	22	22	t718	1	0	t718	YORK& HUM	HO
12673_8_73	2013	22	22	t032	1	0	t032	N WEST	HO
12673_8_74	2013	97	97	t267	1	0	t267	LONDON	HO
12593_2_87	2013	22	22	t492	1	0	t492	S EAST	HO
12593_2_88	2012	22	22	t032	1	0	t032	YORK& HUM	CO
12593_2_89	2013	22	22	t1802	1	0	t1802	YORK& HUM	CO
12593_2_90	2013	30	36	t012	1	0	t012	YORK& HUM	HO
12593_2_86	2013	22	22	t022	1	0	t022	YORK& HUM	CO