The molecular epidemiology and characteristics of methicillin resistant *Staphylococcus aureus* sequence type 22 in a local, national and international context.

By:
Calum Bryson Thomson

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Institute of Microbiology and Infection
College of Medical and Dental Sciences
The University of Birmingham
Edgbaston, Birmingham, UK
January 2017

Abstract

The globally distributed, ciprofloxacin resistant hospital associated methicillin resistant *Staphylococcus aureus* (MRSA) lineage epidemic (E)MRSA-15 (sequence type 22) continues to expand into new healthcare systems around the world. To further understand the evolution of EMRSA-15, which has been proposed to have emerged in the West Midlands, a large collection of contemporaneous and historical Birmingham isolates were studied alongside a collection of International genomes. Through generation of phylogenetic trees this study has been able to determine that EMRSA-15 emerged in c1984 and that during this time a highly related healthcare associated ciprofloxacin sensitive ST22 population was present in Birmingham. A dominant, geographically restricted EMRSA-15 clone was also identified in Birmingham (designated the BHM clone). Analysis of the impact of fluoroquinolone use on the Birmingham EMRSA-15 phylogeny showed increasing use of fluoroquinolones (the introduction of ofloxacin) coincided with the emergence of the BHM clone. Further evidence of geographically restricted EMRSA-15 clones were found in Guernsey and Hong Kong; demonstrating localized evolution giving rise to distinct island populations. Comparison of the desiccation tolerance of geographically restricted and non-restricted EMRSA-15 clones indicated enhanced environmental survival as a possible reason for the spread and dominance of the restricted lineages.

Acknowledgements

Firstly I would like to thank my supervisors Professor Peter Hawkey and Dr Katie Hardy as without their continuing help, support and patience I would not have been able to submit this thesis. I especially thank Dr Matthew Holden (University of St Andrews) for all his help and guidance with my bioinformatics and Bayesian analysis implementation and interpretation. I would also like to thank the Surgical Reconstruction and Microbiology Research Centre (SRMRC) for funding my PhD studies and Vikki Harrison for her administrative and moral support over my four year PhD.

I thank Dr Beryl Oppenheim, Dr Natasha Natnaraja and Dr Margaret Ip for providing me with isolates for my study and Dr Katie Hardy again for letting me use her extensively sampled PhD isolate collections. I thank Dr Nick Loman for allowing me to use his sequencers and to Josh Quick for training me in library preparation. I also thank Sahida Shabir for helping me with all my laboratory work and training, Tim Walker for his help with my VNTR and WGS comparison, Ashley Robinson for his help starting to teach me bioinformatics and Emily Richardson for her bioinformatics expertise and help setting up CLIMB. I additionally thank Sue Manzoor, Claire McMurray, Steve Wilson and Hannah Tanner from HEFT and Grace Richmond, Jess Blair, Michelle Buckner and Ed Bevan from ARG for their help over the years, without whose support my PhD would have been a much more difficult task.

Finally I would like to thank my parents George and Helen for their reassuring support during my PhD as well as my friends Josh, Brian, Toni, River, Rob, Daryan, Anthony and Lee for being there for a strong drink whenever one was needed.

Contents

Table of Figures

Table of Tables

Abbreviations List

Abbreviation	Meaning
3GC	Third generation cephalosporin
A and E	Accident and Emergency
AUC	Area under the curve
BEAST	Bayesian evolutionary analysis sampling trees
BHH	Birmingham Heartlands Hospital
BHI	Brain heart infusion (broth)
BHM	Birmingham Clade
bp	Base pair
BSA	Bovin serum albumin
BSAC	British Society for Antimicrobial Chemotherapy
BWA	Burrows Wheeler Aligner
CA-MRSA	Community associated methicillin resistant *Staphylococcus aureus*
$CaCl_2$	Calcium chloride
CBA	Columbia blood agar
CGE	Centre for Genome Epidemiology
CHB	City Hospital Birmingham
DDD	Defined daily doses
DNA	Deoxyribonucleic acid
EMRSA	Epidemic methicillin resistant *Staphylococcus aureus*
ENA	European Nucleotide Archive
ENT	Ear Nose and Throat (ward)
ESPAUR	English surveillance programme for antimicrobial utilisation and resistance
ESS	Effective sample size
GHH	Good Hope Hospital
GTR	Generalised time reversible model
h	Hours
HA-MRSA	Hospital associated methicillin resistant *Staphylococcus aureus*
HEFT	Heart of England Foundation Trust
IPC	Infection Prevention and Control
ITOL	Interactive tree of life
ITU	Intensive Treatment Unit
LV	Locus variant
MCC	Maximum clade credibility
MCMC	Markov chain Monte Carlo
mg	Miligram
MGE	Mobile genetic element
MIC	Minimum inhibitory concentration
ML	Maximum likelihood
MLST	Multi locus sequence typing
MRSA	Methicillin resistant *Staphylococcus aureus*
NGS	Next generation sequencing

OBD	Occupied bed days
OD	Optical density
PBP2	Penicillin binding protein 2
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEH	Princess Elizabeth Hospital (Guernsey)
PFGE	Pulsed Field Gel Electrophoresis
PHE	Public Health England
PII	Period of increased incidence
QEHB	Queen Elizabeth Hospital Birmingham
QRDR	Quinolone resistance determining region
RAxML	Random axelerated maximum likelihood
RH	Relative humidity
RPM	Revolutions per minute
SD	Standard deviation
SIRU	Staphylococcal Interspersed Repeating Unit
SNP	Single nucleotide polymorphism
ST	Sequence Type
STRD	Summed tandem repeat difference
UK	United Kingdom
USA	United States of America
UV	Ultraviolet
VNTR	Variable number tandem repeat
WGS	Whole genome sequencing
XML	Extensible markup language
µl	Microlitre
σB	Sigma factor B

1 Introduction

1.1 *Staphylococcus aureus*

Staphylococci are Gram positive, non-motile, non spore forming facultative anaerobic cocci that can be seen as pairs or clusters under a microscope which were first identified by Sir Alexander Ogston from a surgical wound abscess in 1881. Shortly following this identification Friedrich Rosenbach demonstrated the organism was responsible for wound infection and named it *Staphylococcus aureus* due to the golden colour of its bacterial colonies (Rosenbach, 1884). Today it is understood that *S. aureus* is a globally distributed opportunistic pathogen that can colonize a range of mammalian host species including humans (Ward *et al.*, 2014, Fitzgerald, 2014). Through the acquisition of resistance to antimicrobial drug classes such as beta-lactams, cephalosporins and fluoroquinolones it has gradually evolved into a pathogen that can cause a high degree of morbidity and mortality as well as becoming a significant economic burden to control.

1.2 Colonization and carriage

S. aureus colonization is distinct from infection, where colonization is the asymptomatic carriage of *S.aureus* in healthy individuals and infection is the invasion of sterile sites that leads to pathology and disease.

The primary niche for human colonization is the anterior nares (Williams, 1963) but *S. aureus* is also carried on the skin in the axilla and perineum, throat and the uro-genital tract of women (Kluytmans *et al.*, 1997, Mertz *et al.*, 2009, Albrecht *et al.*, 2015) with reports of high carriage rates in the throat (Mertz *et al.*, 2009). Human colonization by *S. aureus* has been well defined into persistent and intermittent carriage states with the use of longitudinal screening studies. Up to 20% of the population are persistently

colonized with *S. aureus* and an additional 60% are intermittent carriers who go through periods of being colonized over time. A third carriage state, non carriage, has been proposed previously however this is difficult to prove as screening studies would have to continuously screen individuals for extremely long periods of time or individuals may simply have *S. aureus* at levels below current detection methods (Miller *et al.*, 2014, van Belkum *et al.*, 2009).

Carriage of *S. aureus* is a known risk factor for subsequent disease development (Bode *et al.*, 2010, Edmond and Wenzel, 2013, Kluytmans *et al.*, 1997) and infections are more common in carriers than non-carriers, however mortality risk is known to be higher in patients who newly acquire *S. aureus*. Defined population groups are considered to be at a higher risk of developing disease due to high carriage rates. These include military recruits (Aiello *et al.*, 2006, Zinderman *et al.*, 2004), prison inmates (Aiello *et al.*, 2006, Pan *et al.*, 2003), injecting drug users (Huang *et al.*, 2008, Atkinson *et al.*, 2009, Cooke *et al.*, 2010, Bassetti and Battegay, 2004), athletes (Champion *et al.*, 2014, Karanika *et al.*, 2016) and the homeless (Cooke *et al.*, 2008, Landers *et al.*, 2009). Underlying medical conditions also pre-dispose for *S. aureus* carriage such as human immunodeficiency virus (HIV), diabetes and heart diseases as well as individuals with continuing medical care such as haemodyalisis (Tong *et al.*, 2015b).

1.2.1 Infection

S. aureus infection is the result of the invasion of bacteria into sterile sites that can cause pathology and disease leading to a wide range of infections that range in severity. This process is multifactorial, involving both patient host and bacterial virulence factors to

establish infection such as the loss of the protective skin barrier in injured patients and the multitude of toxins and adhesins produced by *S. aureus* (Otto, 2014). Benign infections are largely superficial skin infection such as impetigo, folliculitis and furunculosis that are relatively common but can include more severe pathologies such as post-operative wound infection. Invasive infections are more severe and often associated with life threatening complications such as endocarditis, osteomyelitis and pneumonia that may require hospitalization for extended periods of time. In addition to specific site infections *S. aureus* is a leading cause of hospital associated bacteraemia, a life threatening condition that is often a sign of secondary site infections, such as infective endocarditis, and can cause a high degree of morbidity and mortality (Tong *et al.*, 2015b). The combination of carriage and scope of disease burden that is coupled with the continued development of antimicrobial resistance has led *S. aureus* to become a significant nosocomial pathogen and a leading cause of hospital acquired infection around the globe.

1.3 Methicillin resistant *Staphylococcus aureus* (MRSA)

The evolution of antimicrobial drug resistance in *S. aureus* has led to the emergence and subsequent global spread of Methicillin Resistant *Staphylococcus aureus* (MRSA); a hospital-associated pathogen that causes a similar array of infections to methicillin sensitive *Staphylococcus aureus* (MSSA) but has resistance to an extended range of antibiotics. This not only reduces the possible treatment options available for serious MRSA infection but infection with MRSA has been shown to have poorer therapeutic outcomes compared to MSSA infection (Cosgrove *et al.*, 2003) leading to increased

morbidity and mortality in the hospital inpatient population and an increased length of hospitalization (Cosgrove *et al.*, 2005, Grundmann *et al.*, 2006).

1.3.1 The evolution of MRSA

In the early 1940s, shortly after the discovery of antibiotics, the first beta lactam antibiotic called penicillin was introduced as a treatment for severe staphylococcal infections which at the time had a mortality rate of approximately 80% (Skinner and Keefer, 1941). In the same year as the first clinical success, penicillinase producing *S. aureus* bacteria were described shortly followed by increasing clinical failures to penicillin (Jevons *et al.*, 1963).

In order to combat this antimicrobial resistance the first and second generation beta lactam penicillins were developed, including methicillin and oxacillin, which were semi synthetic penicillins that were stable to penicillinase (Simon, 1962). Methicillin was introduced into clinical practice in 1959 and two years after this introduction the first clinical failure of methicillin for *S. aureus* treatment was described (Jevons *et al.*, 1963). Subsequent work identified resistance to be due to the presence of a modified penicillin binding protein (PBP2) that was not inhibited by methicillin and was encoded by the resistance gene *mecA* (Enright, 2003) which was carried on the mobile genetic element (MGE) staphylococcal cassette chromosome *mecA* (SCC*mec*) (Katayama *et al.*, 2000). MRSA was seen to cause a similar range of infections as MSSA and remained relatively uncommon after it was first reported (Jevons *et al.*, 1963) and was not seen as a major public health threat.

By 1967 evidence of increasing numbers of multidrug resistant MRSA were being reported across Europe, Australia and India (Jessen *et al.*, 1969, Benner and Kayser, 1968, Jevons *et al.*, 1963, Grundmann *et al.*, 2006). These outbreaks belonged to a single phage type 80A complex that had been widely disseminated. Shortly following this rapid spreading event, MRSA rates began to decline again. The MRSA isolation rate in Switzerland dropped from 20% to 3% between 1971 and 1975 while the frequency of MRSA in Denmark dropped to 0.2% by 1984 (Kayser, 1975, Rosendal *et al.*, 1976). These drastic reductions are thought to have been mediated by changes in antibiotic prescribing and infection control policy being implemented at this time.

In the 1980s MRSA began to develop as a public health concern again as gentamicin resistant clones began to be reported from several countries including the USA (Boyce *et al.*, 1983) and the UK (Duckworth *et al.*, 1988). During this time the first epidemic MRSA (EMRSA) strains were being reported (Kerr *et al.*, 1990, Duckworth *et al.*, 1988) which are defined as a strain that has been identified in two or more patients in two or more hospitals. The first EMRSA strain, EMRSA-1, was identified in London and the South East of England in the 1980s (Duckworth *et al.*, 1988). Since then an additional sixteen EMRSA strains have emerged, with the latest strain being EMRSA-17 however this classification system of MRSA has since been superseded by spa and MLST nomenclature (section 1.4) (Mackintosh *et al.*, 1991, Kerr *et al.*, 1990).

From the early 1990s onwards there was a rapid increase in MRSA hospital associated infections in the UK where MRSA isolation rate increased from 1-3% to 40% (Lowy, 1998, Reacher *et al.*, 2000) but also in Europe and the Americas (Diekema *et al.*, 2001), causing MRSA to become a significant global public health threat requiring immediate intervention to reduce the increasing MRSA infection rate.

Currently the global MRSA rate is on the rise however endemic MRSA prevalence varies between countries (Grundmann *et al.*, 2006, Stefani *et al.*, 2012). As a general rule, the prevalence of MRSA has decline in European countries over the past 10 years, with the Scandanavian countries having the lowest rates of MRSA (Grundmann *et al.*, 2010). The majority of the global burden of MRSA is currently found across Asia, Africa and South American continents which encompass many developing countries. Unfortunately limited data on MRSA prevalence and endemic lineages are available for these countries due to lack of resources to fund extensive public health networks. Varying reports of MRSA rate from different regions have been published however high MRSA rates have been reported in Sri Lanka, South Korea, Vietnam, Taiwan and Hong Kong in Asia (Song *et al.*, 2011, Chen and Huang, 2014) and from Argentina, Chile, Colombia, Guatemala, Mexico, Peru and Uruguay in South America (Guzmán-Blanco *et al.*, 2009).

1.3.2 Infection prevention and control of MRSA

In order to combat the public health threat posed by the increasing rate of MRSA related deaths in hospitals in the late 1990s and early 2000s many countries began to implement infection prevention and control (IPC) policies in hospitals. These policies vary between countries and relate to the background MRSA prevalence and healthcare funding for each country. Countries that have a low MRSA prevalence, The Netherlands and Scandinavian countries (Stefani *et al.*, 2012), tend to adopt "Search and Destroy" methods were MRSA carriers and high risk patients are isolated and barrier nursed by healthcare workers followed by disinfection of rooms after patient discharge. This has been deemed to be successful in reducing both financial costs and patient morbidity (van Rijen and Kluytmans, 2014, Souverein *et al.*, 2016). In higher prevalence countries,

such as Greece or Portugal (Stefani *et al.*, 2012), surveillance policies are employed to suppress MRSA rates to manageable levels (Coia *et al.,* Calfee *et al.*, 2014).

The UK has introduced a stream of IPC policies since 2001, starting with the mandatory reporting of MRSA bacteraemias in 2001 (**Error! Reference source not found.**), which have been extremely successful in continually reducing the MRSA bacteraemia rate. Record low rates of annual MRSA bacteraemia have been achieved annually from 2001 to 2015 with the lowest rate of 750 bacteriaemias per year being reported (Figure 1) (Public Health England, 2016). These policies include recommendations to manage MRSA transmission such as active surveillance, isolation and cohorting colonized or infected patients, decolonization of admitted patients, environmental decolonization and antimicrobial stewardship. These protocols can be employed to different effect however hand hygiene remains the most important measure for preventing all HA infections including MRSA (World Healthcare Organisation, 2009).

Year	Intervention	Methodology
2001	Mandatory MRSA bacteraemia reporting	All MRSA bacteraemias in hospitals reported to UK government
2004	"Clean your hands"	Surveillance of orthopaedic surgical site infections. Mandated placement of bedside alcohol rub. NHS trusts appoint DIPC (director of infection prevention and control)
2004	Targets introduced	Hospitals must aim to meet specific targets (for each hospital) in reduction of MRSA bacteraemias
2005	Cleanliness improvement - Saving lives document	National target reduction for number of MRSA BS infections by 50% over 3 years (Apr 2005 to March 2008) compared to the 2003/4 baseline.
2006	Revised national guidelines	Department of Health (DoH) visits to acute hospitals implemented. CEOs personally responsible for accuracy of infection data for their trusts.
2007	Deep clean	Introduction of bare below the elbow guidance.
2007	Antimicrobial prescribing guidelines	Reduction in use of fluoroquinolones and 3rd generation cephalosporins (3GCs)
2008	Screening elective admission	"Patient safety first". High risk patients screened on admission to selective wards
2010	Universal screening	All patients admitted to a ward are screened for MRSA (nasal/perinneal)
2011	MSSA surveillance	Surveillance and reporting of MSSA blood stream infections
2011	Antimicrobial Stewardship	Hospitals to introduce antimicrobial stewardship policy to control antimicrobial use
2013	PIRs introduced	Post infection reviews of all MRSA blood stream infections

Table 1: Overview of UK national infection prevention and control policy by date of introduction.

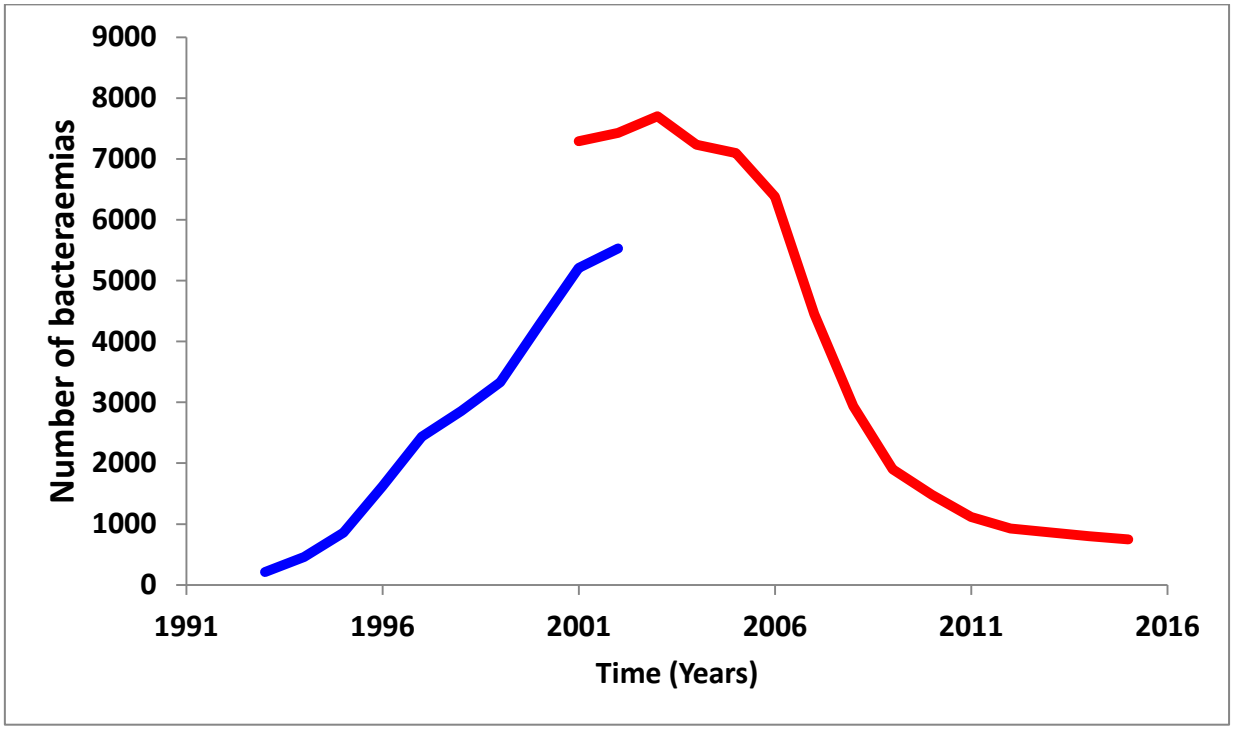

Figure 1: Annual MRSA bacteraemias in England and Wales (UK) from 1993 to 2015. Data before 2001 are from voluntary reports (blue) and after 2001 from mandatory reports (red).

1.3.3 Community associated MRSA

In 1993 a novel MRSA strain from indigenous Australian individuals who had not been exposed to the Australian health care system was identified as the first true community associated MRSA (CA-MRSA) strain (Udo *et al.*). CA-MRSA strains are transmitted in the community, outside the hospital setting in individuals who have no hospital admission or associated risk factors, and differ from HA-MRSA lineages in several ways. Firstly CA-MRSA strains are more susceptible to beta lactam antibiotics compared to HA-MRSA strains; secondly CA-MRSA strains are genotypically different from HA-MRSA strains due to their differing SCC*mec* cassettes and clonal complex (CC) lineages (Section 1.4.3). Finally community isolates are more likely to encode the Panton Valentine leukocidin (PVL) virulence determinant which is often lacking in HA-MRSA strains (Vandenesch *et al.*, 2003, Okuma *et al.*, 2002, Ito *et al.*, 2004). CA-MRSA lineages have also started to be implicated in hospital outbreaks, adding more public health implications to infection control procedures (David *et al.*, 2014).

1.4 Epidemiological typing of MRSA; investigating evolution and transmission

Understanding the genetic relatedness, clonal spread and evolution of MRSA is necessary to facilitate the control of outbreaks and monitor the spread of drug resistant organisms in both local and global settings. Both phenotypic and genotypic methods can be employed to determine the epidemiology of MRSA, however genotypic methods are now largely favored over phenotypic methods due to the higher level of discriminatory power they provide. A variety of genotyping methods are used for delineating MRSA

transmission and epidemiology; many of which exploit genetic polymorphisms in the bacterial genome (Ochman and Davalos, 2006). These techniques have been developed over the past three decades and vary in cost, analysis time and discriminatory power (Table 2), with a general trend of increasing levels of discrimination as new techniques are developed. As these techniques have been developed over time the nomenclature defining MRSA lineages has also with the implementation of new typing methods. MLST is currently the preferred system for categorizing MRSA strains and is described in section 1.4.3.

1.4.1 Phenotypic typing

Phenotypic methods used for typing of *S. aureus* are antibiotic resistance profiles, or antibiograms, and phage typing. Antibiograms consist of a profile of an organism's susceptibility or resistance to a panel of antibiotics (Andrews, 2001) and are standardized (EUCAST guidelines 2016 (European Comittee on Antimicrobial Susceptibility Testing (EUCAST), 2016)), easy to perform and can be automated with the use of robot technology such as the VITEK system (BioMerieux, Marcy-L'Etoile, France).

Phage typing was first developed in the 1940s and involves interpreting the lytic reactions of bacteria to lysis inducing bacteriophage (Wentworth, 1963).This technique has been used in hospitals since the 1960s and is still carried out in healthcare resource poor countries (Williams and Dean, 1974, Wiśniewska *et al.*, 2012, Kali *et al.*, 2013) but is now highly dated and is both time consuming and technically demanding, with limited standardization (Marples and Rosdahl, 1997) and no portability of results. Due to the limited discriminatory power of phenotypic typing methods, genotyping methods are preferred for epidemiological investigation in developed countries.

Typing technique	Equipment and reagent cost	Average cost per isolate	Routine clinical use	Time to results (days)*	Analsis demand	Data portability	Resolution detail
PFGE	Moderate	£4-7	Yes	2-3	Low	Low	Moderate (within lineages)
MLST	High	£20	Yes (reference lab)	2-3	Moderate	High	Low (lineage)
spa	High	£3-5	Yes (reference lab)	1.5	Moderate	High	Low (lineage)
VNTR	Low	£10	Yes (reference lab)	1.5	Low	High	Moderate (within lineages)
WGS	High	~£50$^{\alpha}$	No$^{\beta}$	2-5$^{\gamma}$	High	High	High (SNP level)

Table 2: Comparison of *Staphylococcus aureus* genotyping techniques. * Time inclusive of 24 hour culture, α = Includes culture and DNA extraction costs, β = Used for public health investigations. γ = Dependent on sequencing platform and level of analysis.

1.4.2 Pulsed field gel electrophoresis (PFGE)

PFGE is a discriminatory method for determining if MRSA isolates are clonal and is often considered to be the gold standard in hospital outbreak and transmission investigation. Chromosomal DNA is digested with the enzyme *smaI* to produce 12 or more high molecular weight fragments which are then separated on an agarose gel by electrophoresis using an alternating voltage gradient. Banding patterns produced from the electrophoresis are then interpreted to determine how related isolates are; a process which can be enhanced with the use of software packages such as BioNumberics (Applied Maths, Sint-Martens Latem, Belgium) that can apply unweighted pair group matching or Dice comparisons to improve validity (Golding *et al.*, 2015, Reed *et al.*, 2007). PFGE has several limitations and is both labour intensive and time consuming. All DNA extraction, protein removal and restriction digestion processes involved in PFGE have to be performed in agarose plugs due to PFGE results relying on accurate sizing of high molecular weight DNA fragments which are cleaved easily. Gel electrophoresis requires a minimum of 24 hours for size separation and the use of oscillating frequency gel tanks with pH and temperature controlled buffers (Golding *et al.*, 2015, Reed *et al.*, 2007). Results are reliant on human interpretation and banding patterns can vary between different protocols and laboratory technicians. Furthermore changes in a single restriction site can cause more than one band change, reducing the accuracy of results. This makes PFGE largely un-portable due to results being entirely reliant on banding pattern interpretation that can vary between different diagnostic laboratories (Murchan *et al.*, 2003). Efforts to improve Interpretation have been implemented and validated (Tenover *et al.*, 1995) as well as to standardize PFGE on national and international levels.

The USA and the Netherlands have had success at standardizing PFGE results and reporting on national scales, however there is still no successful programmes for international standardization of PFGE protocols and results.

1.4.3 Multi locus sequence typing (MLST)

MLST is currently the most widely used genotypic method for assigning MRSA into lineages or sequence types (ST) based on single nucleotide polymorphisms (SNPs) occurring in seven housekeeping genes (*arcC, aroE, glpF, gmk, pta, tpi and yqi*) each of which are approximately 500 base pairs (bp) in length. These housekeeping genes are amplified by PCR and then Sanger sequenced, the results of which are compared on an online database. Sequences are compared to all known deposited sequences in a database and assigned a numerical allelic profile that corresponds to a designated sequence type number (Enright *et al.*, 2000). For example the allelic profile 7 6 1 5 8 8 6 corresponds to ST22, profile 3 3 1 1 4 4 3 corresponds to ST8 and profile 2 2 2 2 3 3 2 corresponds to ST36.

Current MRSA nomenclature and population structure definition is based on MLST profiling and often in conjunction with SCC*mec* typing, for example EMRSA-15 is defined as a ST22-SCCmecIV lineage. Highly related MLST lineages are assigned to clonal complexes (CCs) using the Based Upon Related Sequence Types (BURST) algorithm which groups MLST types to a single CC when five of the seven housekeeping genes are identical, therefore creating a group or complex of related lineages. Examples of this

hierarchical structuring of MRSA lineages are detailed in Table 3.

MLST is a useful technique for looking at the global phylogeny of isolates or determining shifts in MRSA populations but has limited use in defining microevolution within hospitals and cannot distinguish between closely related isolates. MLST is also often considered to be labour intensive and time consuming, due to the need for high quality DNA extraction, several rounds of PCR amplification and clean up along with sequencing and post sequencing analysis. The application of next generation sequencing (NGS) methodologies for determining MLST has now reduced both financial and time costs and is discussed in section 1.5.

1.4.4 *spa* locus typing

spa typing is the single locus sequencing of the polymorphic X region of the highly diverse *spa* gene (Protein A) that was developed by Frénay et al in 1996 (Frénay *et al.*, 1996). The diversity of the *spa* gene is due to variation in a number of 24bp repeats within the gene that arise through point mutations, duplications and deletions which changes the number of repeats within the polymorphic X region. Counting the number of repeats can be used to infer a digital profile which can allow different *S. aureus* strains to be determined and compared to each other (Shopsin *et al.*, 1999).

Clonal complex	MLST	spa*	EMRSA	SCC*mec*	Additional Names
CC1	ST1	t128	-	IV	USA400, Canadian MRSA-7
	ST772	t657	-	V	Bengal Bay clone, WA MRSA-60
CC5	ST5	t001	EMRSA-3	II	USA100, New York-Japan Clone, Canadian MRSA-2
	ST228	t041	-	I	South German Epidemic Strain, Italian Clone
CC8	ST8	t008	-	IV	USA300, WA MRSA-12, Canadian MRSA-10
	ST8	t190	EMRSA-2, -6, -12, -13, -14	IV	Lyon Clone
	ST239±	t037	EMRSA-1, -4, -7, -9, -11	III	Czech/Vienna/Hungarian/Portguese/Brazilian Clone, Canadian MRSA-3,-6
CC22	ST22	t032	EMRSA-15	IV	Barnim Epidemic Clone, Canadian MRSA-8
CC30	ST30	t019	-	IV	USA1100, Southwest Pacific Clone
	ST36	t018	EMRSA-16	II	Canadian MRSA-4

Table 3: Example nomenclature of globally distributed MRSA lineages. * = Represents most common spa type. ± = ST239 has 80% homology with ST8 and 20% with ST30.

Single locus sequencing is cheaper than MLST and *spa* typing is highly concordant with MLST. Furthermore publically available centralized *spa* data servers (http://spaserver.ridom.de) (www.seqnet.org) allow *spa* types from over 51 countries to be compared and comprise of over 3900 unique *spa* types and remain one of the largest typing databases for *S. aureus* (Deurenberg and Stobberingh, 2008). WGS methodologies have now been employed to attempt to gain *spa* typing data from next generation sequencing analysis, with a good backwards compatibility being achieved between the traditional Sanger methodology and NGS (Bartels *et al.*, 2014, Bletz *et al.*, 2015).

spa-typing has been used to determine hospital outbreaks and the molecular evolution of MRSA over time (Garvey *et al.*, 2016) and remains the typing method of choice for the UK staphylococcal reference unit (Colindale, Public Health England, UK) however the technique has several limitations. Firstly it cannot differentiate between closely related isolates within the same lineage, a caveat of many typing techniques, and secondly it may not infer evolutionary relationships appropriately due to only analyzing a single locus.

1.4.5 Variable Number Tandem Repeat (VNTR) analysis

Staphylococcal Interspersed Repeating Unit (SIRU) (Hardy *et al.*, 2004) is a *S. aureus* specific VNTR typing scheme that provides high levels of discrimination between two or more isolates by determining the number of consecutively arranged, near identical repeated DNA sequences occurring at seven mini-satellite regions around the genome. These repeat regions are flanked by conserved sequences that are targets for primers and can allow the repeat regions to be amplified by PCR. Amplified products are then separated and sized by gel electrophoresis, which in turn allows a number of repeats to

be calculated for each locus. This generates a digital profile for each isolate, creating a highly portable and high level differentiating typing method (Hardy *et al.*, 2006b, Hardy *et al.*, 2004).

VNTR has been successfully applied to hospital outbreaks (Hardy *et al.*, 2006b), is a cheap and effective method for studying the micro-evolutionary dynamics of hospital MRSA strains and provides a greater level of discrimination than both PFGE and MLST.

1.5 Whole Genome Sequencing (WGS)

Whole genome sequencing (WGS) is the construction of the complete nucleotide sequence of an organisms genome. In 1995 the genome of *Haemophilus influenzae*, the smallest known genome at the time, was fully sequenced and annotated; making it the first bacteria to be whole genome sequenced using Sanger sequencing (Fleischmann *et al.*, 1995). This huge effort took years to undertake and required a six figure budget to complete and is often considered to be "First Generation" whole genome sequencing. In the subsequent 20 years next generation sequencing (NGS), often simply called WGS, has advanced rapidly to become the most appropriate tool for whole genome studies. NGS is the high throughput massively parallelized DNA sequencing of millions of DNA strands to give read sequences, small strands of sequenced DNA, which can be combined to give the nucleotide sequence of the entire genome, therefore drastically increases sequencing capacity to reduce time and cost to sequence multiple genomes (in contrast to capillary electrophoresis Sanger sequencing). NGS technology has rapidly advanced in the past decade in terms of reduced cost and turnaround time, where NGS sequencing capacity

has doubled every 6 to 9 months and now outstrips Moore's law of computer technology whereby computer processor speeds double approximately every 18 months (Loman *et al.*, 2012, Metzker, 2005).

These advances in technology combined with reducing costs have allowed WGS to become a much more accessible tool for investigating all facets of microbiology. In particular WGS has been demonstrated as a highly accurate and discriminatory molecular typing tool for several clinical pathogens due to the ability to provide detail at the single nucleotide polymorphism (SNP) level across an entire genome (Price *et al.*, 2014, Price *et al.*, 2013, Walker *et al.*, 2013, Gardy *et al.*, 2011, Gardy *et al.*, 2015, Koser *et al.*, 2012b, Koser *et al.*, 2012a, Eyre *et al.*, 2012) (Fitzpatrick *et al.*, 2016). Many genotyping techniques such as PFGE and VNTR have been demonstrated to show a reduced ability to discriminate within HA-MRSA lineages, however WGS can scrutinize a considerably larger proportion of the genome than these methodologies. This provides WGS with a much higher discriminatory power than any previous typing scheme which is investigated in this thesis.

WGS has now been used to infer the genetic relatedness of isolates not only in localized hospital settings but also on global population scale (Harris *et al.*, 2010, Price *et al.*, 2013, Fitzpatrick *et al.*, 2016). Furthermore, MLST, *spa* and antibiotic resistance data can all be acquired from the WGS data, providing additional depth of detail being derived from a single technique.

1.5.1 Next generation sequencing platforms

As mentioned above NGS and WGS are terms used for a group of several different NGS sequencing technologies and platforms that are marketed by different industries, each with differing costs and completion times as well as different sequencing chemistry and read lengths produced (Table 4). Of the differing characteristics that define each sequencing platform read length is considered the most important due to smaller reads being more likely to be reproduced in a sequencing run and being more difficult to map to a reference. Increasing read length provides more genetic information, therefore allowing greater overlap between reads and more accurate mapping to the reference genome and a greater reliability of the correct nucleotide at each position. Longer reads also allow sequencing across repetitive regions of the genome which are particularly difficult to map to due to smaller reads mapping in the middle of a repetitive region giving no information on the repetitive region length of position in the genome.

Of the nine currently most used sequencing platforms the Illumina MiSeq and Life Technologies IonTorrent are the first to be used for clinical applications. However, the recently developed Oxford Nanopore minION has significant potential in primary and secondary care as well as field deployment in military hospitals due to it being highly portable with limited power requirements (powered by a laptop) and rapid production of interpretable data.

Currently WGS is still hampered by relatively high costs combined with the requirement for intensive bioinformatics input to make the data understandable and relevant. However continuing reduction in cost and increased automation of analysis,

reducing the need for extensive bioinformatics analysis, will improve the usability of WGS in the clinical setting where it is likely to become a key technique in the future.

It is also prudent to assess prepared libraries for quality and quantity before being applied to sequencers. Several platforms are available including fluorometers (Qubit [Life Technologies]), Fragment size analysers (Bioanalyzer [Agilent Technologies]) or real time quantitative PCR using commercially available kits. qPCR is the most sensitive method for quantifying presence of adaptor tagged fragments due to the specificity of the PCR probes to the adaptor tags on the DNA fragment, while the Bioanalyzer system is most effective at assessing the quality of libraries by determining the range of library fragment sizes.

Manufacturer	Sequencer name	Chemistry	Read Length (base)*	Run time*	Advantages	Disadvantages
Roche	454 GS FLX	Pyrosequencing	700-800	23h	Long read lengths, short run time	Labour intensive, high reagent costs
	GS Junior system	Pyrosequencing	500	8h	Long read lengths, short run time	Labour intensive, high reagent costs
Illumina	HiSeq 2000	Reversible terminator	2x150	11d/2d$^{\alpha}$	Cost effective, high throughput, increeasing read length	Very long run time, short read lengths
	MiSeq	Reversible terminator	2 x159	27h	Cost effective, short run times, widely used.	Shorter read lengths
Life Technologies	ABI/5500xl SOLiD	Ligation	75+35	8d	Low error and high throughput	Very short read lengths, long run times
	IonTorrent	Proton detection	200-400	4h	Short run times. Simple machinery	High error rate, high cost per Mb
	Proton	Proton detection	125	4h	Low cost instrumentation.	High error rate. Increased hands on time.
Pacific Biosciences	PacBio RS	Single molecule synthesis	>10kb	2d	Long read lengths. Simple sample preparation, low reagent cost	High error rate. Very expensive. Difficult installation.
Oxford Nanopore	minION	Solid state nanopore	>510kb	Real-time	Extremely long read length, can sequence across difficult regions. Simple library prep. Highly portable.	Not fully validated at time of writing, high error rate.

Table 4: Next generation sequencing technologies and platforms. Bench top sequencing platforms are coloured blue. α = mode dependent (regular versus rapid mode). * Technologies continually improving. Increasing read length and reduced run time being developed.

1.5.1.1 Illumina sequencing and solexa chemistry

Each of the different available sequencing platforms use different processes to generate NGS read data. Illumina technology employs sequencing by synthesis (SBS) using Solexa chemistry to determine single bases as they are incorporated into a growing sequence read across a large number of DNA template molecules simultaneously. This process is outlined below as an example due the Illumina MiSeq (Illumina, San Diego, USA) being one of the most frequently used technologies which has also been employed in this thesis.

Prepared libraries are denatured to single strand DNA molecules then immobilized on a solid surface flow cell by having one of the sequencing adaptors on the end of the DNA fragment bind to complimentary oligonucleotides bound to the surface of the Illumina flow cell. These oligonucleotides are used as primers to form an initial copy of the bound DNA which is then denatured and the original bound (and now copied) DNA is washed away. The unbound adaptor at the opposite end of the fixed DNA molecule then hybridizes with a bound complementary oligonucleotide on the flow cell to create a bridge structure in a process known as bridge amplification or simultaneous solid phase amplification. Polymerases then form a complementary strand to create a double stranded bridge that is denatured to result in two single stranded copies of the original tagged DNA molecule. This process is repeated simultaneously with other bound DNA fragments to generate clusters of up to 1000 identical copies of each DNA template fragment in highly dense regions across the flow cell.

Sequencing begins with the extension of the first sequencing primer and in each sequencing cycle a mixture of fluorescently labeled deoxynucleoside triphosphates (dNTPs), that have different emission wavelengths for each nucleotide (A C T G), are perfused over the flow cell. A single labeled nucleotide is added to the sequencing read, which is complementary to the template, and after the addition of each nucleotide the fluorescent label is imaged to identify the base. The label is then enzymatically cleaved to allow incorporation of the next nucleotide and the process continues which is known as sequencing by synthesis. This process occurs hundreds of thousands of times in parallel across all clusters on the flow cell, thereby allowing a large amount of DNA material to be sequenced in a relatively short space of time (Illumina Incorporate, 2015a, Illumina Incorporate, 2015b, Loman *et al.*, 2012, Bentley *et al.*, 2008).

1.6 Analyzing WGS data and the need for Bioinformatics

Whole genome sequencing produces millions to billions of reads per run, increasing considerably if high throughput sequencers are being employed, which can take approximately 1GB per *S. aureus* sequence to store (but is dependent on sequencing platform). Therefore any sequencing project requires a good information technology infrastructure to track, analyze and store the WGS data. In addition to high computational demand a significant level of bioinformatics analysis is required to turn raw WGS data into high quality, relevant and interpretable data.

Bioinformatics technology has advanced at a similar speed to sequencing and there is now a multitude of programmes, pipelines and analysis software to manipulate WGS data for a variety of functions. The high number of programmes available from

the Wellcome Trust Sanger Institute, Galaxy Project and Broad Institute alone offer hundreds of different software product options each with differing functionalities meaning the selection of programmes to analyse data can be a daunting task. The exact methodology used to analyze sequencing data is dependent on the length of reads obtained from sequencing, the availability of good reference genomes and the intended application of the data.

For epidemiology and phylogenetic studies an alignment and mapping approach is the most suitable as it allows sequenced reads to be mapped to an annotated reference genome in order to determine SNPs occurring across the genome. These SNPs can in turn be used to infer the relatedness of organisms to each other and generate phylogenetic trees.

1.6.1 Data quality, reference mapping and SNP determination

The quality of WGS produced reads need to be examined as a first step of any WGS analysis. Erroneously identified bases, poor quality sequences or contamination from adaptors can all poorly affect results generated downstream (such as SNP determination) (Bolger *et al.*, 2014, Patel and Jain, 2012). Standard procedures of cleaning data involve adaptor removal and trimming of low quality bases at the end of reads or removing reads from unwanted species (Bao *et al.*, 2014).

Following trimming reads are mapped to a reference genome. Careful consideration should be given when choosing an appropriate reference genome. For example investigating an ST22 MRSA set of isolates one should map to the ST22 reference

genome (HO 5096 0412) as this will give good mapping across the reference genome, therefore maximizing data to determine SNPs from.

Different methods for mapping should be employed depending on the length of read produced by sequencers. For short reads (below 1000bp) programmes that use Burrows-Wheeler Transformation methods are appropriate for rapidly mapping reads to references genomes with limited error (Bao *et al.*, 2014, Li and Durbin, 2010).

Having suitably mapped reads to a reference genome, single nucleotide polymorphisms (SNPs) and insertions or deletions (indels) can now be determined. This is achieved by comparing mapped reads to the reference sequence, with SNPs being identified where bases in mapped reads do not have the same nucleotide as the reference. For bacterial epidemiological analysis this involves determining and comparing the presence or absence of SNPs across all genomes being investigated. SAMtools is one of the most commonly used variant callers which examines every position in the mapped genome and calculates the likelihood that each of these genotypes is truly present in the analysed genome. This can be combined with bcftools which can call SNPs from the genotype likelihoods generated by SAMtools using a defined set of cut-offs to call or reject SNPs (Li *et al.*, 2009, Bao *et al.*, 2014).

Mapped genomes can be viewed with the reference genome in visualization software such as Artemis (Rutherford *et al.*, 2000). This highly useful tool allows users to see where reads are mapping to the genome, review quality of reads, identify gaps in the mapping and see positions of any SNPs that may have been identified.

1.6.2 Phylogenetics and Bayesian Inference

Phylogenetics is the study of evolutionary relationships between biological organisms. As bacteria evolve away from a common ancestor over time they accumulate genetic differences in their genomes in the form of SNPs, insertions or deletions (indels) or the acquisition of mobile genetic elements (MGEs). By comparing genomic data of these phenomenon it is possible to infer the phylogenetic relationship of organisms to each other, usually in the form of a phylogenetic tree (Croucher *et al.*, 2013, Pearson *et al.*, 2009).

WGS provides extensive genomic data for application to phylogenetics, however what specific data to include or exclude can affect the topology and structuring of phylogenetic trees. The "core genome" of bacteria has been defined as the conserved genes that are present in all members of bacterial species. The core genome is interspersed with genes from the "accessory genome" which may be present or absent in individual strains or isolates. The core genome comprises approximately 80% of the *S. aureus* genome with the remaining 20% being the accessory genome which comprises of MGEs such as the SCC*mec* element, phages, plasmids or pathogenicity islands (Lindsay, 2014). SNPs occurring in the core genome are most often separated and retained from those in the accessory genome in order to obtain a reliable scaffold of SNPs to generate phylogenetic trees from. This is because inclusion of SNPs from MGEs can result in misleading phylogenetic interpretations. A caveat of removing MGEs is that it omits parts of the genome that can distinguish isolates and may be influencing micro-evolutionary events (Planet *et al.*, 2016), however the repertoire of MGEs in a population can be determined using *de novo* assembly.

Reliable SNP information can be use to display the relatedness of isolates to each other in the form of a phylogenetic tree. This tree contains nodes connected by branches, with each node representing the divergence of isolates away from a common ancestor (Yang and Rannala, 2012). Different methods for phylogenetic reconstruction are available, the most commonly published of which are neighbor joining (NJ), maximum likelihood (ML) or Bayesian maximum clade credibility trees (MCC). Neighbor joining is the most basic algorithm for tree generation and uses a distance method to construct trees based on pairwise distance (SNP distance) between isolates. Maximum likelihood is the estimation of unknown parameters in a model (Felsenstein, 1981) and compares all sequence sites in an alignment to build trees with more evolutionary relevant branch lengths. Bayesian inference is a method of statistical inference where parameters in a model are considered to be random variables with statistical distributions. This methodology was applied to phylogenetics (Bayesian phylogenetics) in the late 1990s (Yang and Rannala, 2012) and has the benefit of allowing variable mutation rate across branches in the phylogenetic tree (Drummond and Rambaut, 2007). Bayesian phylogenetics is however considerably more complex to perform than ML or NJ techniques but allows time-linked MCC trees and can demonstrate at what point in time isolates diverged from each other.

The choice of which type of tree to produce is therefore dependent on the question being asked. For example NJ tree is appropriate for determining number of SNP relating isolates in an outbreak where MCC trees can identify when a new strain or clone emerged. The differing models of tree construction also have differing process

times, with NJ trees being more rapid to construct compared to ML trees and is also dependent on the number of genomes in the analyses.

1.6.3 Application of WGS in MRSA epidemiological studies

1.6.3.1 Hospital transmission and infection control

WGS is highly applicable for hospital transmission investigation and infection prevention control for several key nosocomial pathogens. This is due to the high level of discriminatory power provided from determining SNPs across the core genome that cannot be achieved with other typing techniques. This high discriminatory power can however become difficult to interpret when used to determine transmission between patients. If a large number of SNPs distinguish two isolates (over 100 SNPs) from different patients it is clear that there is significant genetic distance between the isolates, indicating they are unrelated and a transmission event can be reliably excluded. However when a smaller number of SNPs distinguish isolates a transmission event is less easy to exclude. Defined cut offs for excluding transmission events have been suggested at 2.5 SNPs (Fitzpatrick *et al.*, 2016) or 3 SNPs (Salipante *et al.*, 2015) in *Acinetobacter baumanii*, 2 SNPs in *Clostridium difficile* (Eyre *et al.*, 2013, Didelot *et al.*, 2012) or 12 SNPs for recent transmission of *M. tuberculosis* (Walker *et al.*, 2013). In *S. aureus* cut offs have been suggested at greater than 23 SNPs to exclude a transmission event (Uhlemann *et al.*, 2014) while other authors suggest <40 SNPs (Golubchik *et al.*, 2013). This is due to the cloud of variation that has been demonstrated in MRSA carriers, where different colonies from a single nasal swab from an individual have differed from each other by up to 40 SNPs (Young *et al.*, 2012,

Golubchik *et al.*, 2013) (Harris *et al.*, 2013). This phenomenon has been demonstrated in *A. baumanii* and *C. difficile* (Eyre *et al.*, 2013, Fitzpatrick *et al.*, 2016) suggesting caution should be used when using SNPs to identify transmission events and ought to be used in conjunction with epidemiological data where possible.

Despite uncertainty around defining transmission WGS has been used in real time MRSA outbreak situations. Harris *et al* (2013) were able to link two outbreaks across space and time in a neonatal intensive care unit that could not be confirmed by conventional typing methods. WGS was able to determine transmission within the unit, between mothers and to the community and confirm MRSA carriage by a staff member was the reservoir for linking the two outbreak periods.

1.6.3.2 Global epidemiology

WGS has also been used to further understand the global epidemiology and evolution of MRSA. Harris *et al* conducted the first application of WGS for global MRSA epidemiological analysis in 2010 (Harris *et al.*, 2010). The authors were able to show strong evidence of intercontinental spread of ST239 between Asia and the UK as well as endemic sub-clones of the lineage in Thai hospitals. More recent studies have used both MCC and ML trees to date the emergence and global transmission of EMRSA-15. Holden *et al* (2013) used an MCC tree to show the evolution of the ciprofloxacin resistant ST22 strain EMRSA-15 emerged in the West Midlands of the UK in the late 1980s before spreading across the UK and to Europe (Holden *et al.*, 2013) (Section 1.8.1). Further evidence of EMRSA-15 transmission from the UK to Portugal and

Germany has been demonstrated by Aanensen and colleagues who used ML trees with a snapshot population of ST22 across Europe to show single importations to Portugal and Germany from the UK (Aanensen *et al.*, 2016). Hsu *et al* and colleagues also showed the importation of ST22 into the Singaporean healthcare system and tracked its evolution over time in comparison to the previously dominant ST239. Data showed multiple sub-clades of ST239 isolates but a single highly related clade of ST22 isolates, suggesting multiple introductions of ST239 from the surrounding regions, but only a single introduction of ST22 at the turn of the millennium, which was consistent with previous literature (Hsu *et al.*, 2015, Holden *et al.*, 2013).

1.6.3.3 Evolutionary rate

A key factor in understanding phylogenetics is the rate at which mutations occur to accumulate SNPs in the genome which is known as the molecular clock rate. This rate varies between bacteria, with *Mycobacterium tuberculosis* evolving very slowly at 0.5 mutations per genome per year ($1.1x10^{-7}$ mutations per site per year) (Ford *et al.*, 2011) compared to *Helicobacter pylori* at 30.4 mutations per genome per year ($1.9 x10^{-5}$ mutations per site per year) (Kennemann *et al.*, 2011).

There is current debate as to how the molecular clock changes in different environments such as in nasal carriage versus clinical infection or the hospital environment. The clock rate is also unlikely to be fixed and how differing selection pressures that are likely to change with time and therefore affect the evolutionary rate of mutation.

1.7 Global Epidemiology of MRSA

1.7.1 Globally distributed lineages of MRSA

Over the past 30 years the methodologies outlined in section 1.4 have been used to understand the localized, national and global epidemiology of MRSA. It is now established that different lineages are endemic to various regions around the globe.

ST239 is the most widely distributed MRSA lineage and is dominant throughout most of Asia, accounting for regions where 60% of the world's population are resident and MRSA rates are high (Feil *et al.*, 2008, Chongtrakool *et al.*, 2006, Ko *et al.*, 2005). MRSA lineages vary across Europe with different lineages predominating in different countries. ST228 and ST5 are endemic in many European countries such as France, Germany and Italy, however the dominant UK lineage ST22 (see section 1.8.1) has steadily replaced the dominant lineages in these countries and is now the fastest growing HA-MRSA lineage in Europe (Grundmann *et al.*, 2014).

In contrast in the USA the clinical landscape is currently dominated by the CA-MRSA lineage USA300 (ST8, CC8). Studies in healthcare facilities in the USA have shown varying reports of the prevalence of USA300; ranging from 60% to 98% of MRSA isolates being related to USA300 (Deurenberg and Stobberingh, 2008).

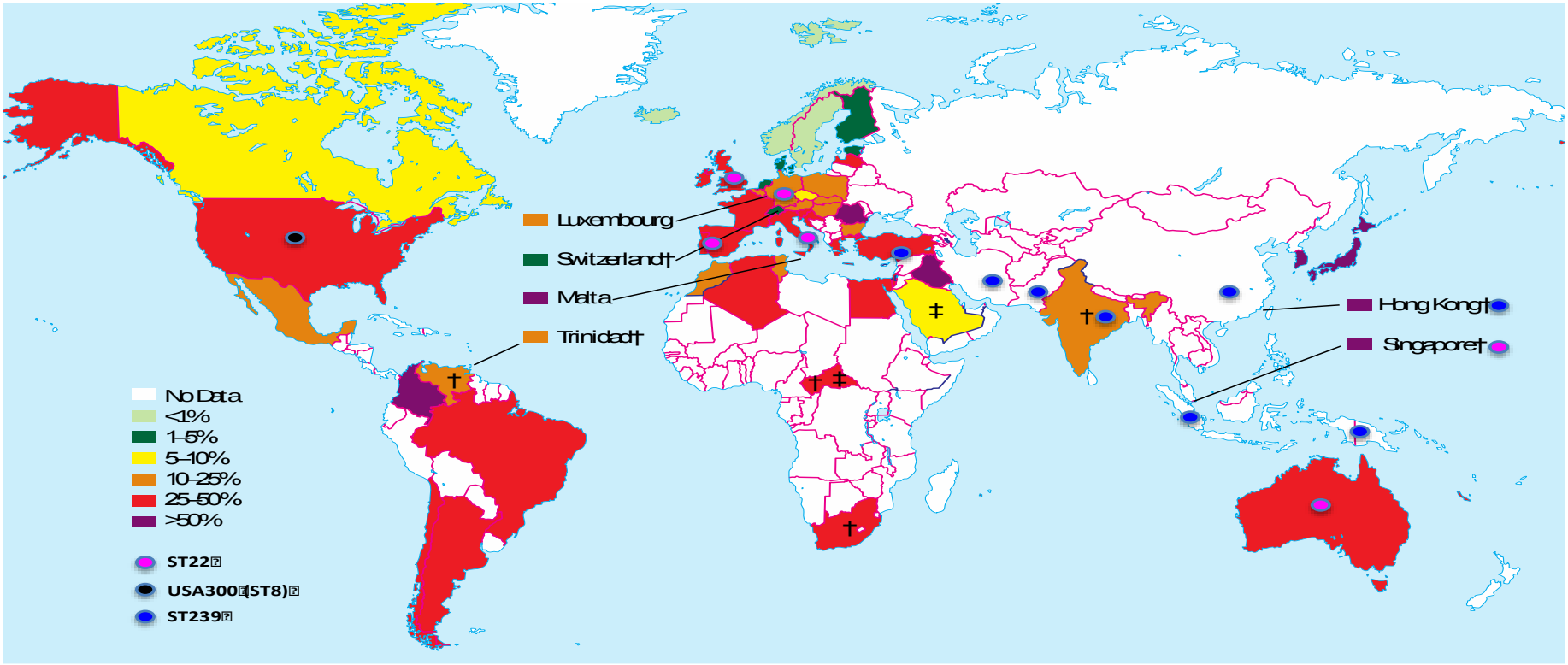

Figure 2: The global prevalence of MRSA. Countries are coloured by MRSA prevalence. Coloured circles represent distribution of three of the most dominant MRSA lineages where data available. =Estimates are based on a study from a single hospital. ‡=Prevalence estimates are based on studies between 1993 and 1997. Adapted from Grundmann *et al* (2006).

1.8 MRSA in the UK

1.8.1 Epidemic methicillin resistant *Staphylococcus aureus* 15 (EMRSA-15); the dominant UK clone.

The current UK MRSA landscape is dominated by the ST22 HA-MRSA strain EMRSA-15. The strain is classically identified by being urease negative, produces enterotoxin C, has weak lysis with phage 80/81 and is resistant to ciprofloxacin. EMRSA-15 isolates are usually PVL negative, as are all UK HA-MRSA isolates, with less than 2% of UK clinical isolates being PVL positive (Teare *et al.*, 2010).

The evolutionary origin of EMRSA-15 was recently described by Holden and colleagues who hypothesized that the strain emerged in the West Midlands (UK) in the late 1980s (Holden *et al.*, 2013). Using a collection of globally sampled ST22 isolates the authors proposed that a ciprofloxacin sensitive ST22 population present in the West Midlands in the 1980s acquired resistance to ciprofloxacin via mutations in the DNA topisomerase genes *gyrA* and *grlA* (Holden *et al.*, 2013, Griggs *et al.*, 2003) (discussed in section 1.9). This ciprofloxacin resistant strain (EMRSA-15) spread rapidly throughout the UK, during which time the national MRSA bacteraemia rate also rapidly increased (Figure 1). In 2001, EMRSA-15 accounted for greater than 60% of UK MRSA bacteraemias (Johnson *et al.*, 2001), the height of the MRSA epidemic in the UK, and rose to 85% in 2007 (Ellington *et al.*, 2010).

The rapid expansion of EMRSA-15 across the UK has been suggested to be facilitated by EMRSA-15 having inherent resistance to ciprofloxacin. This resistance is thought to have provided a fitness advantage allowing the resistant strain to spread rapidly around the

country as ciprofloxacin use increased in hospitals, increasing from 18.9% of total antibiotic usage in 1992 to 31.5% in 1999 (Livermore *et al.*, 2002). Further evidence of the rapid dissemination of EMRSA-15 demonstrated by Reuter *et al* who investigating the phylogeny of MRSA in the UK on a national scale. Using over 700 EMRSA-15 bacteraemia isolates sampled over 10 years the authors were able to show geographical population structuring across the UK which they attribute to the rapid emergence, spread and subsequent local diversification of the clone (Reuter *et al.*, 2015). Evidence for the evolution of EMRSA-15 in the Birmingham region remains somewhat debatable. Limited isolates from the West Midlands or any isolates pre-dating 1990 have been included in previous evolutionary studies. Additionally no study has extensively examined the longitudinal phylogenetic evolution of EMRSA-15 in the region it emerged in, which is addressed in this thesis.

EMRSA-15 is currently a globally disseminated clone, having been exported from the UK to Europe, Australia and localized regions in Asia where it has been shown to repeatedly replace established dominant lineages in both high and low resource healthcare systems (Witte *et al.*, 2001, Knight *et al.*, 2012, Hsu *et al.*, 2007a, Hsu *et al.*, 2015, Dhawan *et al.*, 2015, Coombs *et al.*, 2014, Coombs *et al.*, 2004, Aanensen *et al.*, 2016, Dailey *et al.*, 2005). A recent WGS study by Aanensen *et al* (2016) has shown that EMRSA-15 was exported to Berlin from the UK before spreading outwards throughout Germany and similarly with an importation to Lisbon in Portugal (Aanensen *et al.*, 2016), replacing the dominant clones CC5 (Albrecht *et al.*, 2011) and ST239 (Amorim *et al.*, 2007) respectively. In Singapore ST239 was the dominant MRSA lineage (Ko *et al.*, 2005, Feil *et al.*, 2008) however between 2003 and 2015 there was a notable increase in ST22 MRSA

strains being isolates from clinical specimens in Singapore where EMRSA-15 is now the dominant lineage (Hsu *et al.*, 2005, Hsu *et al.*, 2015, Hsu *et al.*, 2007b). Furthermore other globally dominant lineages such as ST239 have caused sporadic outbreaks in UK hospitals however they repeatedly fail to become established as the dominant clone, suggesting EMRSA-15 is highly adapted to the UK hospital system. The reason for the dominant and highly adaptable nature of EMRSA-15 is not fully understood, however improved stress responses compared to other MRSA lineages and the inherent resistance to fluoroquinolones (Baldan *et al.*, 2015, Baldan *et al.*, 2012, Knight *et al.*, 2012) are thought to play an important role (discussed in section 1.10).

1.9 Resistance to Fluoroquinolone antibiotics in MRSA

1.9.1 Fluoroquinolone antibiotics

Quinolones, and subsequent fluoroquinolones, are a class of synthetic broad-spectrum antimicrobial agents first discovered in the early 1960s that fulfill many of the criteria defining a "good antibiotic drug". Qualities such as high potency, broad spectrum of activity (Gram positive and negative bacteria), high serum levels and availability of oral and intravenous formulations have given strength to the benefits of using quinolone antibiotics to treat infections (Andersson and MacGowan, 2003, Redgrave *et al.*, 2014).

The first quinolones were derived from the naturally occurring quinine and acted by directly inhibiting DNA replication via binding and creating drug-protein complexes with DNA gyrase and topoisomerase IV; two highly conserved proteins that are essential in the supercoiling of bacteria DNA (Drlica, 1990). DNA gyrase is required

for the relief of torsional stress that accumulates in front of DNA replication forks by introducing negative supercoils and is the only known enzyme capable of actively unwinding nucleic acids. Topoisomerase IV is involved in decatenating newly replicated daughter chromosomes. Both proteins achieve their function by covalently binding to DNA to introducing double strand DNA breaks and form enzyme-DNA complexes. These complexes are altered with the addition of a fluoroquinolone which creates a drug-enzyme-DNA complex by forming a water metal ion bridge between oxygen molecules in the amine group of the drug and the hydroxyl group of the enzyme. The enzymes are then no longer able to re-ligate the cleaved DNA causing a build up of cleaved DNA complexes and triggering the SOS response, ultimately leading to cell death (Baranello *et al.*, 2012, Anderson *et al.*, 1999, Redgrave *et al.*, 2014, Aldred *et al.*, 2013).

Fluoroquinolone antibiotics were created to increase the potency of quinolone drugs by replacing the eighth carbon of the molecule "backbone" with a nitrogen atom and the addition of a fluorine molecule at position 6 (Figure 3) (Ball, 2000). This initial modification caused a 10 fold increase in DNA gyrase inhibition and a 100 fold improvement in MIC (Andersson and MacGowan, 2003). Over the years the fluoroquinolone drugs have been further modified in accordance with clinical need; with structural changes resulting in increasing in potency and spectrum of activity (Table 5) and subsequent increasing global widespread use of the drug class.

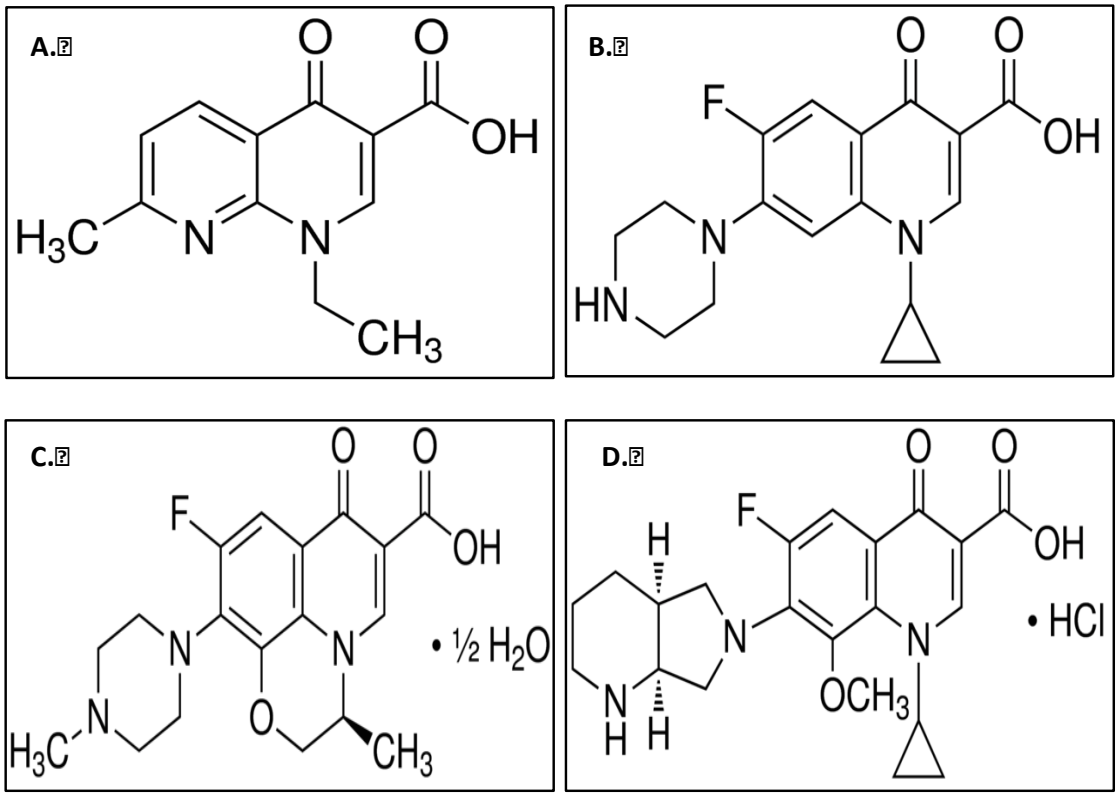

Figure 3: Chemical structure of quinolone and fluoroquinolone drugs: (A) Nalidixic acid, (B) Ciprofloxacin, (C) Levofloxacin hemihydrate and (D) Moxifloxacin. Chemical structures were obtained from the Sigma Aldrich website (www.sigmaaldrich.com).

Year of FDA approval	Drug	Generation	MIC_{90} (*E. coli*)	Primary use/treatment	Note
1967	Nalidixic acid	1st	8	UTI	-
1986	Norfloxacin	2nd	0.12	UTI, Gonorrhoea	-
1987	Ciprofloxacin	2nd	0.03	UTI, tissue infection	-
1990	Ofloxacin	2nd	0.12	Occular infection	-
1995	Sparfloxacin	3rd	0.06	Pneumonia, abdominal infection	Withdrawn from UK Market
1996	Levofloxacin	3rd	0.12	SSTI, UTI, pneumonia	-
1997	Trovafloxacin	4th	0.06	Pneumonia, abdominal infection	Withdrawn from UK Market
1999	Moxifloxacin	4th	0.06	Pneumonia, abdominal infection	Withdrawn from UK Market

Table 5: Fluoroquinolones licensed for clinical use in the UK by year of FDA approval. MIC values obtained from Andersson and McGowan (2003).

1.9.2 Mechanism of fluoroquinolone resistance in *Staphylococcus aureus*

As with many antimicrobial agents, increasing use of fluoroquinolones over the past 30 years had led to increases in antimicrobial resistance and highly resistant isolates of *S. aureus* are now prevalent. Unlike many other resistance mechanisms, resistance to fluoroquinolones is primarily mediated by point mutations in one or more of the genes encoding for DNA gyrase ,*gyrA* and *gyrB*, or topoisomerase IV, *grlA* and *grlB* (in *S. aureus*). Non-synonymous point mutations arising in short DNA sequences known at the quinolone resistance determining region (QRDR) are known to cause changes in fluoroquinolone drug susceptibility. If these mutations cause amino acid changes that alter the structure of either subunit of the two enzymes, the drug target site of the fluoroquinolone can be changed. This reduces the drug binding affinity and can lead to drug resistance (Hooper, 2000, Piddock, 1999, Hooper, 1999, Hooper, 2001). Many of these mutations have been well described in *S. aureus* (Table 6) with some mutations having greater decrease in drug susceptibility than others. In Gram negative bacteria mutations in *gyrA* are thought to be most influential in generating resistance (Takahashi *et al.*, 1998), however the primary drug target in *S. aureus*, and other Gram positive bacteria, is topoisomerase IV (Anderson *et al.*, 1999) and mutations in *gyrA, gyrB* and *grlA* genes are associated with clinical resistance. In resistant strains mutations in *gyrA* are always preceded by mutations in *grlA* and mutations in *gyrA* and *gyrB* alone have been seen to not cause resistance. Conversely mutations in the *grlA* gene with or without *gyrA* mutations have been described in fluoroquinolone resistant strains. Combinations of mutations in *gyrA* and *grlA* have been shown to give increased minimum inhibitory concentrations (MIC) compared to mutations in *grlA* alone. An example of this is the

combination of an S80F mutation in *GrlA* and a S84L mutation in *gyrA* giving a 128-fold increase in ciprofloxacin resistance compared to wild type strains (Griggs *et al.*, 2003, Horii *et al.*, 2003). Interestingly mutations in *grlB* are less widely reported than mutations in *grlA* or *gyrA* but appear to have a minimal impact on fluoroquinolone resistance in combination with mutations in *grlA* and *gyrA* (Takahashi *et al.*, 1998).

In addition to the alteration in drug target described above, fluoroquinolone resistance can also arise from drug efflux. The membrane efflux pump NorA, encoded by the gene *norA*, has been shown to cause increased resistance to hydrophilic fluroquinolones including norfloxacin by actively pumping the drug out of the bacterial cell and decreasing intracellular concentration. Increased expression or mutations arising in the coding region of the *norA* gene have been shown to increase resistance to fluoroquinolones (Kaatz *et al.*, 1993, Kaatz *et al.*, 1999, Yoshida *et al.*, 1990).

Enzyme	Gene	Mutation	Reference
Gyrase	*gyrA*	Ser-84-Lys	(Griggs *et al.*, 2003)
	gyrA	Gly-88-Leu	(Takahashi *et al.*, 1998)
	gyrB	Asp-437-Asn	(Ito *et al.*, 1994)
	gyrB	Arg-458-Gln	(Ito *et al.*, 1994)
Topoisomerase	*grlA*	Ser-80-Phe	(Griggs *et al.*, 2003)
	grlA	Ala-116-Glu	(Hooper, 2000)
	grlB	Arg-470-Asp	(Fournier and Hooper, 1998)
	grlB	Pro-451-Ser	(Hannachi-M'Zali *et al.*, 2002, Horii *et al.*, 2003)

Table 6: Characterized mutations in fluoroquinolone target genes.

1.9.3 Fluoroquinolone use and MRSA evolution

As mentioned previously fluoroquinolone drugs are used across the globe in human and veterinary medicine due to their high potency and spectrum of activity. This frequent use over the past 25 years has led to an increase in fluoroquinolone resistance in species such as *Escherichia coli*, *Klebsiella pneumoniae* and *S. aureus* across Europe, with clinical MRSA isolates resistant to fluoroquinolones now being widespread (Livermore *et al.*, 2002).

In the UK, changes in prescription policy have been implemented to reduce the use of fluoroquinolones in primary and secondary care, partially due to the link between fluoroquinolone use and *Clostridium difficile* infection (Deshpande *et al.*, 2015). Recent data from the English surveillance programme for antimicrobial usage and resistance (ESPAUR) report has shown that there has been a slight reduction in the use of fluoroquinolnes in England and Wales between 2010 and 2013, however this trend has been reversed with an increase in use from 2014 (Public Health England, 2015).

While not a first line drug for staphylococcal infections fluoroquinolone resistance is thought to be a biomarker for successful lineages, with major successful HA-MRSA clones being resistant. These include EMRSA-15 (Richardson and Watson, 2013) (detailed in section 1.8.1), EMRSA-16 (McAdam *et al.*, 2012), ST 239 (Lindsay, 2013) and the CA-MRSA strain USA 300 (Alam *et al.*, 2015). The widespread fluoroquinolone resistance in MRSA likely due to fluoroquinolones being readily secreted in sweat through body regions such as the axilla and perineum which are colonization sites for MRSA (Hawkey, 1997). Furthermore the selective advantage

resistance would provide in the hospital environment where fluoroquinolone drugs would be frequently used provides reasoning for resistance in HA-MRSA lineages.

Due to the large number of fluoroquinolone resistant isolates emerging in a range of bacterial species, the evolutionary fitness advantage of fluoroquinolone resistance has been widely debated. Alterations in supercoiling genes are thought to have global cellular responses, with changes to supercoiling shown to increase expression of over 100 genes and decrease expression of 200 genes in *E. coli* (Peter *et al.*, 2004). Evolutionary experiments with *E. coli* have suggested mutations within supercoiling genes are also under selection for fitness in the absence of antibiotic pressures (Crozat *et al.*, 2005). Studies determining the relative fitness of mutations in supercoiling genes in *E. coli* and *S. pneumoniae* have shown mutations to result in a fitness cost; with mutant strains having poorer survival in co-culture and growth rate experiments (Rozen *et al.*, 2007, Marcusson *et al.*, 2009). On the other hand mutations in *gyrA* in *Campylobacter jejuni* (Han *et al.*, 2012) and *Salmonella enterica* (Baker *et al.*, 2013) have been shown to give a fitness gain, allowing strains to become fluoroquinolone resistant and out compete wild type strains in drug free competition experiments.

It has previously been demonstrated that a fitness advantage is required for dissemination of MRSA clones in hospitals (Thouverez *et al.*, 2003) however experiments directly relating to flouroquinolone resistance and fitness in *S. aureus* are limited. Horvath *et al* have shown that different MRSA lineages suffer different fitness costs upon developing resistance to fluoroquinolones (Horváth *et al.*, 2012). For example an ST239 isolate had reduced fitness compared to an ST5 isolate when both isolates had the same MIC value; which the authors attribute to the change in dominant lineage in their region over time (the replacement of ST239 with ST5). The authors also

demonstrated that higher MIC values cause a greater loss of fitness, with isolates with increasing MIC values have reduced fitness. Interestingly the data also showed that EMRSA-15 isolates had the highest fitness, with an EMRSA-15 isolate having comparable fitness to an ST5 isolate with a two fold lower MIC to ciprofloxacin and an ST239 isolate with a 64 fold lower MIC to ciprofloxacin (section 1.8.1).

While limited molecular studies have investigated fitness advantage of fluoroquinolone use in MRSA, several studies have investigated the impact of fluoroquinolone use on the MRSA isolation rate in hospitals. Exposure to fluoroquinolones has been shown give an increase in MRSA isolation rate (Weber *et al.*, 2003) and a systematic review has implicated fluoroquinolones are the antibiotic class having most impact on MRSA infection incidence in hospitals (Tacconelli *et al.*, 2008). Data has shown that increasing use of fluoroquinolones in hospitals has been linked with increasing MRSA rate (Lafaurie *et al.*, 2012, Knight *et al.*, 2012, Weber *et al.*, 2003) and decreasing use has induced a decrease in MRSA rate (Lafaurie *et al.*, 2012). Re-introduction of fluoroquinolones into a hospital after period of restricted use has also been shown to cause a significant increase in MRSA isolation rate to pre-restricted use (Parienti *et al.*, 2011).

A caveat of studies linking fluoroquinolone use to MRSA prevalence is they often fail to include factors other than antimicrobial prescription in their MRSA impact analyses. Several studies have investigated the usage of alcohol based hand rubs (ABHR) in litres per year and have shown that increasing use of ABHRs led to a decreasing trend of MRSA isolation (Lafaurie *et al.*, 2012, Parienti *et al.*, 2011). However most studies fail to account for important infection prevention and control (IPC) interventions such as

elective or universal screening of inpatients, decolonization procedures or patient isolation. A recent and unique study by Lawes *et al* (Lawes *et al.*, 2015) has attempted to address these issues by assessing the impact of both antibiotic stewardship and IPC policies on MRSA prevalence across a health board in Scotland over a period of 16 years. In 2008 Scotland developed it's first national antibiotic stewardship plan (Nathwani *et al.*, 2011) to reduce use of the 4C antibiotics (clindamycin, ciprofloxacin [and other fluoroquinolones], cephalosporins and co-amoxiclav) and macrolide antibiotics. Comparisons of antibiotic prescribing data, MRSA rate and IPC interventions showed restricted use of antibiotics reduced hospital MRSA prevalence by 46% but that IPC procedures also had important contributions to recent reductions in MRSA burden. However co-linearities in associations between the different classes of antimicrobials analysed meant that the MRSA reduction could not be directly linked to reduction in fluoroquinolone drug use and the exclusion of data for staffing levels, patient transfer rates and compliance with isolation and decolonization protocols due to inconsistency further demonstrates the difficulty of quasi-experimental studies.

Fluoroquinolone use has therefore been implicated in the continuing evolution of MRSA. While there is evidence of the impact fluoroquinolone use has in generating resistance and how it can affect hospital MRSA isolation rate, no study has investigated the impact clinical fluoroquinolone use has on detailed MRSA phylogenetic evolutionary structures which is investigated in this thesis.

1.10 Desiccation Tolerance in *Staphylococcus aureus*

The environment plays an important role in MRSA outbreaks and control due to it being a potential MRSA contamination reservoir and can therefore transmit to patients and lead to infection (Hardy *et al.*, 2006a, Boyce, 2007). Several hospital surfaces or inanimate objects are readily colonized by *S. aureus* including mattresses, curtains, door handles, beds, telephones, patient notes and medical equipment (Boyce, 2016). In order for an environmental surface to act as a reservoir the bacteria must be able to survive on that surface or media. Several hospital pathogens are capable of surviving in the hospital environment, most notably *Clostridium difficile*, *Acinetobacter baumanii* (Jawad *et al.*, 1996) and *S. aureus* (Boyce, 2007).

In the natural environment bacteria are exposed to many stresses such as starvation, exposure to UV or short chain fatty acids for those that colonize skin. One of the major stresses bacteria face is desiccation; the process of extreme drying where bacteria lose cellular water causing them to become damaged and die. The loss of water causes the cell capsular layer to shrink, increases intracellular salt and macromolecule concentrations (Nocker *et al.*, 2012, Potts, 1994) and further damage is caused, particularly to DNA and proteins, by the increase in concentration of cellular oxygen radicals. These radicals are produced from the Fenton reaction where hydrogen peroxide is reduced to form highly reactive hydroxyl species that increase in the desiccated cell. Gram positive bacteria are considered to be more resistant to desiccation, partially due to their lack of LPS or flagella and thicker peptidoglycan walls which reduce net loss of water (Nocker *et al.*, 2012).

1.10.1 Experimental methods for desiccation assays/tolerance

Noble (Noble, 1962) and Rountree (Rountree, 1963) were two of the first authors to investigate the desiccation tolerance of *S. aureus* in the early 1960s. Their findings showed *S. aureus* to be able to survive for an extended period time on hospital fabrics and that different *S. aureus* strains had different tolerances to desiccation. Since then many other studies have been performed on the desiccation tolerance of *S. aureus* and MRSA that are detailed in Table 7. These studies are difficult to compare directly due to the variability in their methodology, with studies using different strains, inoculum concentrations, substrate growth materials or recovery media. The duration of study is also highly variable between studies and to date the longest duration MRSA strains have been analysed is 3 years, with 0.001% of bacteria being able to survive for this duration (Chaibenjawong and Foster, 2010). Furthermore it is almost impossible to measure desiccation tolerance without including other stress factors such as starvation and UV exposure. Despite these variable parameters most studies come to similar conclusions regarding several key aspects to desiccation tolerance which are outlined in the sections below.

Author (reference)	Publish date	Strains (number)	Innoculum concentration	Relative Humidity (%)*	Temperature (°C)	Protectants	Test Material	Max duration of experiment	Survival time
Noble (Noble, 1962)	1962	MSSA (5+)	-	-	-	-	Cotton blanket	-	>203d
Rountree (Rountree, 1963)	1963	MSSA, MSSA clinical and screening (19 total)	1.0 $x10^7$ to 2.0 $x10^7$	42-50	19-27	None.	Cotton	90d	>15d
Pettit and Lowbury (Pettit and Lowbury, 1968)	1968	MSSA (3)	2 $x10^7$	35-50	25	None	Glass	24h	>24h
Lacey (Lacey *et al.*, 1970)	1970	MSSA (38)	$1.0x10^3$ to $1.0x10^4$	40-60	25, 37	None	Glass	12h	>12h
Lacey (Lacey, 1972)	1972	MRSA (10)	Not stated	-	30, 36	none	Glass	6h	>6h
Beard-Pegler (Beard-Pegler *et al.*, 1988)	1988	MRSA (16), MSSA (9), CNS (12)	$5.0x10^7$	50-67	22-29	None	Cotton	-	<9w
Duckworth and Jordens (Duckworth and Jordens, 1990)	1990	EMRSA-1 (5), MSSA (13)	7.0 $x10^7$	-	37	None	Formica	14d	>14d
Hirai *et al* (Hirai, 1991)	1991	MRSA, MSSA	1.5 $x10^5$	50	21	None	Cotton	25d	>25d
Farrington *et al*	1992	MRSA, MSSA	-	42-54	22-25	-	Cotton	-	> 79d
Jawad *et al* (Jawad *et al.*, 1996)	1996	MSSA (3)	$2x10^7$	10, 31 and 93*	22	BSA	Glass	35d	>18d
Wagenvoort (Wagenvoort *et al.*, 2000)	2000	MRSA (5)	$1.0x10^9$	20 - 49	-	-	Hospital dust/plastic	409d	>315d
Chaibenjawong *et al* (Chaibenjawong and Foster, 2010)	2011	MRSA, MSSA, E-16, COL, S. aureus mutants (22 total)	Not stated	-	20, 25, 37	BSA, inert proteins, catalases, sugar protectants	Plastic	3y	>3y
Knight *et a*l (Knight *et al.*, 2012)	2012	ST22 [EMRSA-15], ST36 [EMRSA-16]	$1.0x10^6$	-	-	-	Plastic	5d	> 5d
Baldan *et al* (Baldan *et al.*, 2015)	2015	ST22 [EMRSA-15] (10), ST228 (10)	Not stated	-	-	-	plastic	7d	> 7d
Domon *et al (Domon et al., 2016)*	2016	MRSA (5), MSSA (4)	$2\text{-}5x10^6$ and $2\text{-}5x10^4$	-	-	-	Plastic	8d	>8d

Table 7:Desiccation survival studies of methicillin sensitive and resistant *Stapylococcus aureus*. Studies that have controlled relative humidity are denoted by a * in the relative humidity column.

1.10.2 Test environment factors

In 1950 Lidwell and Lowbury first noticed that the daily death rate of *S. aureus* changed with the relative humidity of the environment (Lidwell and Lowbury, 1950), showing that different environmental humidity affected how bacteria survived over time. Since then relatively few studies have actively controlled the relative humidity that test bacteria are exposed to (Jawad *et al.*, 1996), however two have demonstrated natural changes in room temperature and humidity changed the number of bacteria recovered in their experiments (Rountree, 1963, Beard-Pegler *et al.*, 1988).

1.10.3 Test material and suspension media

The material *S. aureus* is inoculated onto determine its desiccation tolerance and therefore has strong implications as to how long it survives. Variable survival rates have been demonstrated on glass (Lacey *et al.*, 1970, Pettit and Lowbury, 1968, Lacey *et al.*, 1986, McDade and Hall, 1964, Jawad *et al.*, 1996), cotton or other hospital fabrics (Rountree, 1963, Beard-Pegler *et al.*, 1988, Hirai, 1991, Noble, 1962) and inert plastics (Duckworth and Jordens, 1990, Knight *et al.*, 2012, Baldan *et al.*, 2015, Wagenvoort *et al.*, 2000, Domon *et al.*, 2016). These test materials reflect the various materials found in the hospital environment, suggesting some hospital surfaces can harbor viable bacteria longer than others. MRSA appears to survive for longer on cotton fabrics compared to glass however to date the longest record of MRSA survival has been three years on dry plastic (Chaibenjawong and Foster, 2010, Neely and Maley, 2000). Interestingly bacteria inoculated onto cotton fabric in experiments by Rountree (Rountree, 1963) and Beard-

pegler *et al* (Beard-Pegler *et al.*, 1988) showed bacterial counts to have increased between primary inoculation and day one (24 hour timepoint). This may be due to the environmental temperature and humidity in the laboratory setting of these experiments in which humidity and temperature were not controlled.

The addition of protectants in the form of inert protein has also shown to increase the long term survival of MRSA. Chaibenjawong *et al* noted a 15 fold increase in survival of MRSA strains suspended in 3% bovine serum albumin (BSA), the concentration of protein in "dirty hospital conditions" and a 5 fold increase with the addition of boiled catalase (Kawamura-Sato *et al.*, 2008, Tuladhar *et al.*, 2012), compared to those suspended in water (Chaibenjawong and Foster, 2010) with similar findings being seen in *Acinetobacter* species (Jawad *et al.*, 1996).

1.10.4 Methicillin resistant and desiccation tolerance

Lacey *et al* in 1972 (Lacey, 1972) were the first authors to include MRSA strains in their analysis; finding no difference in survival between methicillin resistant and sensitive strains. This finding has been supported by some literature (Lacey *et al.*, 1986, Hirai, 1991, Chaibenjawong and Foster, 2010) but refuted by others (Duckworth and Jordens, 1990, Domon *et al.*, 2016). Beard-pegler *et al* (Beard-Pegler *et al.*, 1988) and Lacey (Lacey, 1972) showed no relationship between antibiotic resistance to various antimicrobial agents and the desiccation tolerance of different strains based on phage typing and the acquisition and loss of resistance determinants respectively. On the other hand Duckworth *et al* (Duckworth and Jordens, 1990) showed multi-drug

resistant EMRSA-1 was able to survive better than both MSSA and MRSA strains with fewer drug resistances, however this greater survivability may have been due to inherent features of the EMRSA strain rather than it's antimicrobial resistant properties. Finally a recently published article by Domon *et al* (Domon *et al.*, 2016) showed that methicillin sensitive strains survived longer than resistant strains; perhaps due to the lack of the SCC*mec* element in the MSSA strains.

1.10.5 Inoculum concentration

The concentration of bacteria inoculated onto test material has been shown to be important in their length of survival. The majority of MRSA desiccation studies use inoculum concentrations ranging from $1.0x10^7$ to $1.0x10^9$ cfu per ml however Domon *et al* (Domon *et al.*, 2016) showed bacteria inoculated onto propylene tubes at lower concentrations, $1.0x10^4$ cfu per ml compared to $1.0x10^6$ cfu per ml, had poorer survival and faster cell death. The established infective dose for MRSA is $1.0x10^5$ cfu per ml (Schmid-Hempel and Frank, 2007) and the total aerobic bacteria count on human hands is approximately 2.0 to $3.0x10^6$ cfu per ml (McGinley *et al.*, 1988), suggesting lower inoculum concentrations may more accurately represent MRSA transmission by human contact.

1.10.6 MRSA lineages with different desiccation tolerance

Different MRSA lineages have different inherent properties associated with them including their tolerance to desiccation. Sequence type 22 (ST22) has been repeatedly

shown to outcompete other lineages such as ST36 (EMRSA-16) (Knight *et al.*, 2012), which evolved at a similar time to EMRSA-15, or ST228 (Baldan *et al.*, 2015) that is endemic across several countries in Europe. In both cases ST22 had a significantly higher number of bacteria surviving at 120h and 168h post inoculation respectively. EMRSA-15 was also able to outcompete both strains in co-culture, had greater acid and alkali tolerance as well as thermo-tolerance and osmostic stress resistance (Baldan *et al.*, 2015, Baldan *et al.*, 2012). EMRSA-1 strains have also been shown to survive longer than non-epidemic strains, surviving twice as long (14 days) as the MSSA strain "oxford staph" (NCTC 6571) (Duckworth and Jordens, 1990). Studies that have investigated different strains have selected isolates based on antibiotic profiles or phage typing however due to the study pre-dating the MLST typing scheme it is difficult to relate these data to current nomenclature and MRSA lineages.

Only a limited number of studies have examined the differences in desiccation tolerance between well defined MRSA lineages and none have examined desiccation tolerance within an MRSA lineage. As most HA MRSA clones are pandemic, individual STs are likely being exposed to different environmental stresses in different geographical locations. Therefore evolutionary adaptation to these specific environments may have occurred between different clones of a specific ST. To date no study has investigated desiccation tolerance between EMRSA-15 clones from different geographical niches which has been investigated in the work described here.

1.11 Cellular mechanisms for desiccation tolerance in *S. aureus*

S. aureus has several regulation factors that contribute to it's ability to tolerate desiccation. Sigma factor B (σB) is an RNA polymerase global regulator that is known to control over 250 genes (Bischoff *et al.*, 2004) and has been identified in many Gram positive and negative bacteria. Unlike other bacteria, *S. aureus* has a limited repertoire of sigma factors to regulate many different cellular mechanisms (Chaibenjawong and Foster, 2010, Shaw *et al.*, 2008) and σB has been demonstrated to be important in stress responses, particularly in the stationary phase, (Clements and Foster, 1999) including desiccation tolerance. σB mutants have been shown to have reduced tolerance to desiccation (Chaibenjawong and Foster, 2010) which is mediated through a number of mechanisms. Sigma B has also been demonstrated to control katA production (Cosgrove *et al.*, 2007) and staphyloxanthin biosynthesis (Pelz *et al.*, 2005, Clauditz *et al.*, 2006) both of which are involved in desiccation stress tolerance.

The only catalase gene in *S. aureus* is *katA* which produces catalase that is required for tolerance to oxidative stress by reacting with oxygen radical species and desiccation tolerance is increased 12 fold when media is supplemented with catalase (Chaibenjawong and Foster, 2010). Furthermore mutants lacking the catalase production gene *katA* (Cosgrove *et al.*, 2007) have shown reduced survivability over time.

Staphyloxanthin is an orange-red triterpenoid carotenoid pigment that was first described by Marshall and Wilmoth (Marshall and Wilmoth, 1981) which gives *S. aureus* it's golden colour and is coded by the five gene operon *ctrNOPQM* (Pelz *et al.*, 2005). Staphyloxanthin plays important role in desiccation tolerance by absorbing energy from

oxygen radicals (Pelz *et al.*, 2005) and greater killing of bacteria is seen in mutants lacking functional staphyloxanthin (Liu and Nizet, 2009). Furthermore the pigmentation properties of the carotenoid are thought to protect the cell from UV damage in the environment (Potts, 1994).

Recently an additional sigma factor, σS, has been suggested as an alternate factor for regulation and desiccation tolerance. σS mutants showed reduced long term survival, due to desiccation and other environmental stresses, however transcription profiling showed the factor is not expressed under normal growth conditions (Shaw *et al.*, 2008), indicating a secondary role of σS in stress tolerance.

1.12 Hypotheses

- EMRSA-15 emerged in the Birmingham region in the mid 1980s.
- The Evolution of EMRSA-15 in Birmingham has been influenced by the clinical use of fluoroquinolone antibiotics.
- Geographically restricted EMRSA-15 populations will exhibit limited diversity
- EMRSA-15 clones or sub-lineages exhibit differences desiccation tolerance.

1.13 Aims

- Develop a bioinformatics pipeline that will analyse WGS data to determine SNPs occurring in the core genome and construction of phylogenetic trees of EMRSA-15 genomes.

- Use clinical and screening isolates collected between 1985 and 2015 from Birmingham in conjunction with an international collection of genomes to further elucidate the evolutionary origins of EMRSA-15.

- Investigate the phylogenetic structure of EMRSA-15 from geographically restricted regions, its transmission around the globe and its continuing evolution in Birmingham and using the isolate collection from aim 2.

- To apply comprehensive antimicrobial prescription data from a single hospital trust to a phylogenetic tree of Birmingham genomes in order to determine the impact of antimicrobial prescription on the phylogenetic evolution of EMRSA-15.

- Compare the discriminatory power of VNTR and WGS SNP typing methodologies to discriminate isolates and identify transmission events within the Birmingham EMRSA-15 population.

- To establish if desiccation tolerance has contributed to the success of different geographically restricted EMRSA-15 clones by investigating their desiccation resistance in a controlled environment.

2 Materials and Methods

2.1 Media

All *Staphylococcus aureus* isolates used in this study were cultured on pre-poured ISO sensitest agar or Columbia blood agar (CBA) plates provided by Oxoid (Basingstoke, UK).

2.2 Storage of Isolates

All bacterial isolates in this study were stored long term in "Protect microorganism preservation tubes (Technical service consultants Ltd, Heywood, UK) tubes. Each isolate was grown at 37°C for 24 hours on ISO sensitest agar (Oxoid, Basingstoke, UK) and then suspended in the cryotube liquid using a sterile cotton swab. Each preserve tube was labeled with the strain ID and storage date and stored at -80°C. When required, strains were sub-cultured directly from frozen stock onto CBA or ISO sensitest agar and incubated at 37°C.

2.3 Isolate Collections

In order to study the evolution of EMRSA-15 in Birmingham over time a total of 317 isolates from retrospective and prospective sample collections that were sampled from circa 1985 to 2015 were sourced (Table 8). These collections were then screened for EMRSA-15 isolates using VNTR typing (detailed in section 2.5). VNTR typing was used to identify EMRSA-15 isolates is currently the most appropriate tool available, being cheaper, faster and more discriminatory than *spa* typing.

Collection	Date Range	Location (Country)	Number of Isolates VNTR typed	Number of ST22 Isolates*	Number of isolates WGS	Clinical	Screening
HEFT ITU	2002-2004	Birmingham (UK)	36	36	36	0	36
HEFT Surgical	2005-2007	Birmingham (UK)	33	33	33	0	33
HEFT Contemporary	2012-2015	Birmingham (UK)	64	54	54	15	39
QEHB Contemporary	2013-2015	Birmingham (UK)	55	36	36	14	22
CHB Contemporary	2013-2015	Birmingham (UK)	48	18	18	4	14
CHB Historical	1985	Birmingham (UK)	14	1	1	1	0
Guernsey (PEH)	2005-2010	Guernsey (British Channel)	54	21	21	8	13[α]
Hong Kong	2010-2014	Hong Kong (China)	13	13	13	4	9
Total	-	-	317	212	212	46	153

Table 8: Isolate collections used in the study. * Isolates classified as ST22 from VNTR typing result comparison to NCTC EMRSA-15 control. [α] eight unknown screening or clinical isolates.

Example Strain	Locus							Profile
	L1	L13	L15	L16	L21	L5	L7	
EMRSA-15	1	5	0	3	16	-	2	1 5 0 3 16 - 2
EMRSA-16	1	2	2	2	5	3	2	1 2 2 2 5 3 2
MSSA-476	1	6	2	2	11	3	2	1 6 2 2 3 2 2

Table 9: Examples of the seven digit VNTR profile patterns where the number of repeats at each locus is displayed as an integer. Throughout this thesis VNTR profiles are displayed in the order L1, L13, L15, L16, L21, L5 and L7.

Isolates with one, zero and three repeats at VNTR loci L1, L15 and L7 were considered to be ST22 or EMRSA-15 due to previous work identifying EMRSA-15 isolates to demonstrate this profile (Table 9).

2.3.1 Historical Isolate collection

A collection of 14 clinical isolates that had been sampled from City hospital Birmingham (CHB) in circa 1985 were grown on cysteine lactose electrolyte deficient (CLED) media (Oxoid, Basingstoke, UK) to select for *S. aureus* by presence of bright yellow colonies which were confirmed as *S. aureus* by latex agglutination test (Thermo Scientific, Waltham, USA). Only a single isolate from this collection was identified as ST22 by VNTR typing.

2.3.2 Retrospective screening isolate collection

Thirty six isolates from 36 patients were selected from a collection of 477 screening isolates sampled from 117 MRSA positive patients admitted to an intensive treatment unit (ITU) at the Heart of England Foundation Trust (HEFT) between February 2001 and December 2003. During the study period 412 patients were admitted to ITU each patient admitted to ITU was screened on admission and every three days for the duration of their stay in ITU as well as having any infected wounds swabbed.

Thirty three isolates from 33 patients were selected from a second inpatient nasal screening collection sampled between November 2005 and April 2007 from surgical

wards at HEFT. All patients admitted to seven surgical wards (two general surgery wards, trauma and orthopaedics, vascular, ear nose and throat (ENT), thoracic and urology wards) were screened on admission and then every four days until discharge.

For both retrospective collections isolates were selected on matching specific criteria. Due to the collections having been previously VNTR typed, isolates were primarily selected on VNTR typing data. Only isolates corresponding to an EMRSA-15 VNTR profile (number of repeats at each locus) (Table 9) were included, leaving a pool of 198 isolates from 95 patients in the ITU screening collection and to 1098 isolates from 714 patients in the surgical screening collection.

Isolates were then selected to maximize sampling across time (as required for BEAST analysis) with at least one isolate per month being included where possible. Due to ITU study design there were no isolates sampled between October 2002 and June 2006 for this collection.

Isolates selected from the ITU screening study between 2001 and 2003 are hereby designated as the "HEFT ITU" collection and those from surgical wards between 2006 and 2008 are designated as "HEFT surgical" collection (Table 8).

2.3.3 Contemporaneous Birmingham Isolate collections

One hundred and sixty seven contemporaneous isolates from three hospitals in Birmingham (HEFT, City Hospital Birmingham (CHB) and the Queen Elizabeth Hospital Birmingham (QEHB) were collected (Table 8, Figure 4) from routinely saved collections

(Table 8). These comprised of 64, 55 and 48 isolates from HEFT, QEHB and CHB respectively and were all sampled from individual inpatients from surgical, trauma and ITU wards at each of the hospitals.

VNTR typing identified 54, 36 and 18 EMRSA-15 isolates from HEFT, QEHB and CHB respectively to give a total of 108 (33 clinical and 75 screening) isolates.

2.3.4 Isolates from around the World

Fifty four isolates from the only acute hospital on Guernsey (UK) were sourced, 21 of which were designated as EMRSA-15 by VNTR typing. Of these 21 isolates, eight were sampled from an ITU ward over a period of four years between 2006 and 2010. The remaining 13 isolates were sampled from sterile sites across five inpatient wards between 2013 and 2015.

Thirteen previously identified as ST22 isolates from four district general hospitals in Hong Kong (China) were also included. These comprised of eight nasal screening isolates from neonatal and pediatric wards across two hospitals sampled in 2014 and five blood culture isolates from three hospitals sampled between 2010 and 2011.

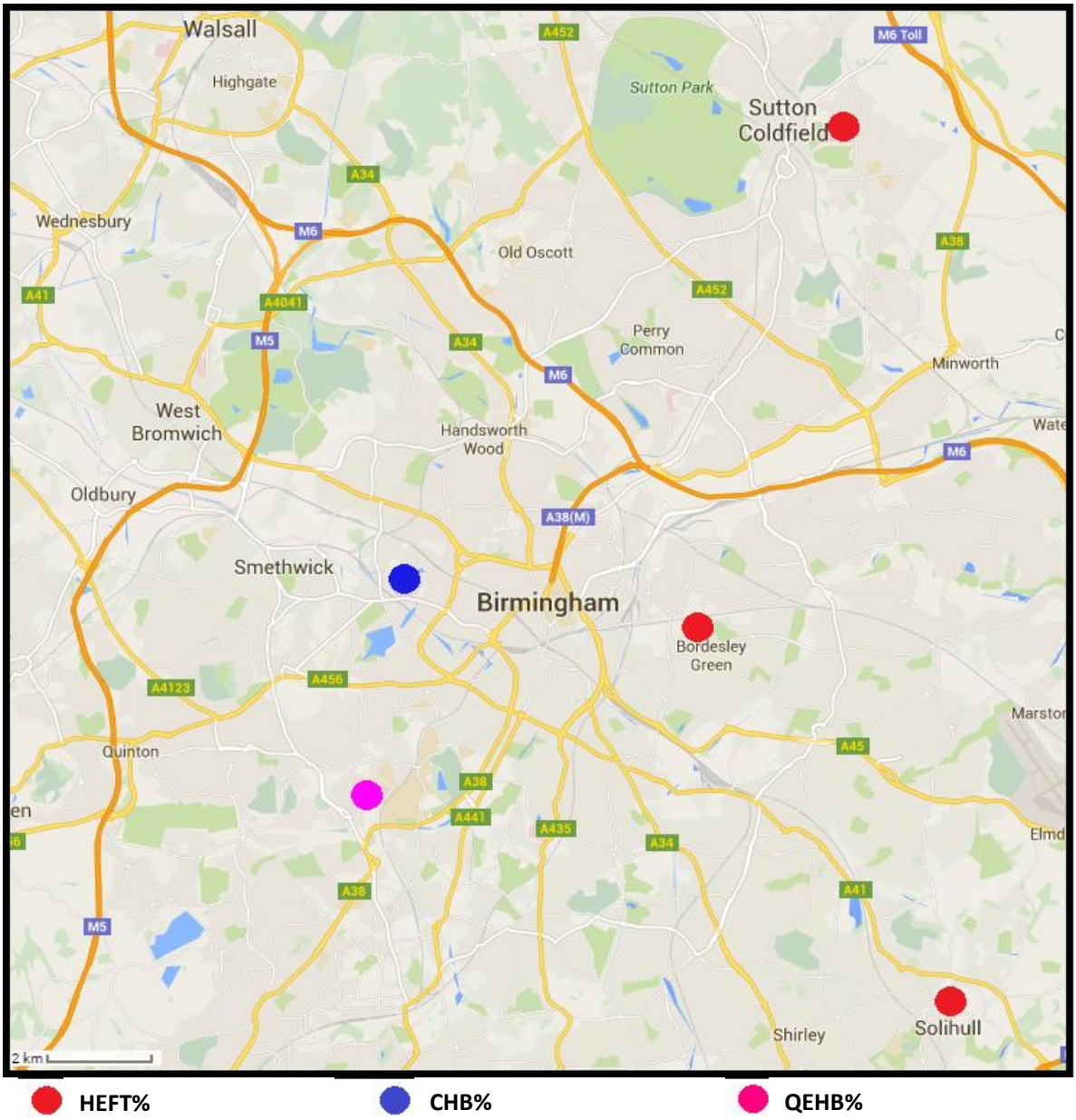

Figure 4: location of sampling hospitals in the Birmingham conurbation.

2.4 Downloaded genomes

In order to contextualize the Birmingham phylogeny on a national and global scale 227 publically available EMRSA-15 genomes were downloaded in fastq format from the European Nucleotide Archive (ENA) (http://www.ebi.ac.uk/ena/) (Table 10).

One hundred and twenty three genomes from the Holden *et al* global ST22 population study (Holden *et al.*, 2013) consisting of 13 ciprofloxacin sensitive ST22 isolates and 110 ciprofloxacin resistance ST22 (EMRSA-15) isolates were downloaded. This collection included the first three reported EMRSA-15 genomes from the West Midlands that were sampled between 1991 and 1992 as well as genomes from Germany, Portugal, the Czech republic, Australasia and Singapore.

An additional one hundred and six EMRSA-15 bacteraemia genomes sampled over a period of 10 years from the UK national Reuter et al (2016) study were also included. This study collected and sequenced a total of 783 EMRSA-15 genomes from 46 laboratories (across 16 UK regions) that submitted MRSA bacteraemia isolates to the British Society for Antimicrobial Chemotherapy (BSAC), collecting between 58 and 92 EMRSA-15 isolates per year.

My selection of isolates from this study included all 73 genomes that were collected from three hospitals (33 from Coventry and Warwickshire, 24 from Royal Shrewsbury and 16 from CHB) across the West Midlands and a single genome from each English submitting region (11 regions excluding the West Midlands) in 2001, 2005 and 2010. Only a single isolate was available from one English region therefore giving 31 genomes selected for analysis. Delineation of submitting hospital was requested from BSAC for the West Midlands hospitals only.

Collection	Locations	Date range	Number of Genomes
Holden (International)	UK	1990-2008	49
	Europe	1997-2008	58
	Australia	1999-2006	10
	Singapore	2004-2007	7
Reuter (UK)	UK (West Midlands)	2001-2010	73*
	UK (England Regions)	2001-2010	30
Total	-	-	227

Table 10: Downloaded genomes included in the study. * Includes 13 isolates from CHB.

2.5 Variable Number Tandem Repeat (VNTR) Staphylococcal interspersed Repeating Unit (SIRU) analysis

DNA was extracted from overnight cultures of MRSA at 37°C on ISO sensitest agar for use as a DNA template in PCR reactions. Three colony picks were suspended in 25µl of 100µg/ml lysostaphin (Sigma Aldrich, St Louis, USA), vortexed and incubated at 37°C for 10 minutes to lyse the cellular membrane. Lysate was then incubated in 25µl of 100µg/ml proteinase K (Sigma Aldrich, St Louis, USA) and 75 µl 0.1M pH 8.0 Tris (Sigma Aldrich, St Louis, USA) at 37°C for 10 minutes followed by 95°C for an addition 10 minutes. Solutions were then centrifuged at 13000 rpm for two minutes to pellet any cell debris and leave DNA extract in solution. DNA extracts were stored at 4°C or at -20°C long term.

PCR mastermixes were prepared in a laminar flow cabinet under PCR clean conditions. Each PCR reaction contained 15.9ml of molecular grade water (Thermo Scientific, Waltham, USA), 1 X PCR buffer (Thermo Scientific, Waltham, USA), 1µM left and right flanking primer (Eurofins, Luxembourg), 200µM of each deoxynucleoside triphosphate (dNTP) (BioGene, Huntingdon, UK), 2mM $MgCl_2$ (Thermo Scientific, Waltham, USA), 0.5U AmpliTaq Gold (Thermo Scientific, Waltham, USA) and two µl DNA extract to give a total volume of 25µl per reaction. Amplification was performed in a block system on a Mastercycler Pro (Eppendorf, Hamburg, Germany). PCR reaction had a pre-denaturation step at 94°C for 2 min, denaturation at 94°C for 30s followed by 35 cycles of annealing at 55°C and extension at 72°C for one minute. A final extension phase at 72°C for two minutes was also included.

PCR amplicons were analysed by capillary gel electrophoresis using the QIAGEN QIAxcel E gene (QIAGEN, Hilden, Germany). A separation programme of OL500 for each of the VNTR loci with the exception of locus 21 which has the separation programmme OM1100. All runs were analysed with a 50bp ladder (Promega, Madison, USA). The number of repeats was calculated from the amplicon size using the repeat unit and flanking region base pair sizes as outlined by Hardy and colleagues 2004 (Hardy *et al.*, 2004).

EMRSA-16 DNA extract was used as a positive control for all VNTR tests due to its amplifying across all loci and a negative control of one µl water instead of DNA extract was also included. Isolates were considered unique if they varied at any of the seven loci.

2.6 Whole Genome Sequencing (WGS)

2.6.1 Qiagen DNA extraction

MRSA isolates were inoculated onto ISO senstiest agar (Oxoid, Basingstoke, UK) from -80^{o}C stored cultures and incubated for 24 hours at 37^{o}C before DNA was extracted from single colony picks using QiaAmp DNA mini kits (QIAGEN, Hilden, Germany).

Three colony picks were suspended in 180µl lysis solution (135µl 0.1M tris-HCl and 45µl of 200µg/ml lysostaphin) in a 1ml sterile Eppendorf tube (Eppendorf, Hamburg, Germany) and incubated at 37°C for 60 minutes to denature the bacterial cell wall. The nuclear envelope was then denatured by adding 20µl proteinase K (Qiagen Hilden, Germany) and 200µl buffer AL(Qiagen Hilden, Germany). Tubes were vortexed

and incubated at 56°C for 30 minutes followed by incubation at 95°C for 15 minutes to denature the proteinase K. Two hundred microliters of absolute ethanol was added to the tube, to elute DNA, which were vortexed for 15 seconds followed by short spin centrifugation to collect droplets. The 600µl volume from the tube was transposed to a Qiagen DNA spin column (Qiagen, Hilden, Germany) and centrifuged at 8000rpm for one minute and collected supernatant was discarded. Five hundred microlites of buffer AW1 (Qiagen, Hilden, Germany), which contains guanidinium chloride for denaturing proteins, was applied to the column which was centrifuged at 8000rpm for one minute and collected supernatant was discarded. Five hundred microlites of buffer AW2 (Qiagen, Hilden, Germany), which contains ethanol and removes salts to purify DNA, was applied to the column which was centrifuged at 14000rpm for three minutes and collected supernatant was discarded. Clean collection tubes were applied to each of the columns and 200µl of buffer AE (10 mM Tris-Cl, 0.5 mM EDTA; pH 9.0) was applied to the column which was then incubated at room temperature for 15 minutes. The column was then centrifuged at 8000rpm for 3 minutes to collect purified DNA which was transferred to a 0.5ml PCR clean Eppendorf and stored at -20°C.

DNA concentrations were verified by fluorometry using the Qubit dsDNA high spec assay kit (Invitrogen, Carlsbad, USA) and the Qubit 2.0 flourometer (Invitrogen, Carlsbad, USA). Quantification of a single DNA sample is as follows: Qubit working stock was made by adding 3µl Qubit dsDNA high spec reagent to 597µl Qubit dsDNA high spec buffer in a 1ml PCR clean Eppendorf to give a volume of 600µl. One hundred and ninety microlitres of working stock was added to each of two PCR clean thin-walled, clear, 0.5ml PCR tubes along with 10µl of each Qubit quantification standard. One hundred and

ninety eight microlitres of working stock solution was then added to PCR clean thin-walled, clear, 0.5ml PCR tube to which 2µl of purified DNA were added. All PCR tubes were vortexed for 10 seconds and incubated at room temperature for two minutes. Tubes containing standards were applied to the fluorometer first, in order of size, to create a standard curve that DNA concentrations could be quantified against before applying the tube with DNA.

Extracted DNA was then diluted to 0.2ng per µl in molecular grade water for Nextera XT sample preparation.

2.6.2 Sample library preparation and MiSeq instrumentation

Samples were prepared for sequencing using Nextera XT DNA Sample Preparation Kit (Illumina, San Diego, USA). For each sample five microliters of DNA at 0.2ng/ul was mixed with 10µl of tagment DNA buffer (Illumina, San Diego, USA) and 5µl of amplicon tagment mix (Illumina, San Diego, USA) in a 96 well PCR clean plate and heated at 55^{o}C for 10 minutes using a thermocycler (Eppendorf, Hamburg, Germany). A separate well in the plate was used for each MRSA isolate DNA extract. Five microliters of neutralize tagment buffer (Illumina, San Diego, Germany) was added to each well and incubated at room temperature for five minutes.

Each tagmented DNA sample was then amplified by a limited cycle PCR reaction and given a unique barcode index to allow sample pooling and cluster formation when being sequencing on the MiSeq. Five microlitres of index 1 and five microlitres of index 2 along

with 15µl of Nextera PCR mastermix were added to each well of the PCR plate. The PCR plate was then put in a thermocycler (Eppendorf, Hamburg, Germany) and run with the following conditions: pre- denaturation step at 72°C for 3 min, denaturation at 95°C for 30s followed by 12 cycles of 95°C for 10 seconds, 55°C for 30 seconds and 72°C for 30 seconds, and a final step of 75°C for five minutes.

DNA libraries were purified using AMPure XP beads (Beckman Coulter, California, USA) to remove short library fragments. Fifty microlitres of each PCR product was transferred to a 96 deep well midi plate (Thermo Scientific, Waltham, USA) containing 30µl of AMPure XP beads. Midi plates were then sealed and shaken at 1800rpm for two minutes using a plate shaker. Following shaking the midi plate was placed on a magnetic stand for two minutes until the liquid in each well had become clear and a visible pellet of beads was evident in each well. Supernatant material was collected and discarded using a 200µl pipette. The midi plate was removed from the magnetic stand and each well was washed with 200µl of 80% ethanol before being placed back on the magnetic stand for 30 seconds and removing the supernatant with a 200µl pipette. This wash procedure was repeated twice then wells were allowed to air dry for 15 minutes while on the magnetic stand. After 15 minutes the plate was removed from the magnetic stand and 52.5µl of re-suspension buffer (Illumina, San Diego, USA) was added to each well and shaken on a plate shaker at 1800rpm for two minutes. Plates were left to incubate at room temperature for two minutes before being placed on a magnetic stand for two minutes or until the liquid was clear with an evident pellet at the bottom of the well. Fifty microlitres of supernatant from each well was aliquoted into in a new PCR clean midi plate using a 200µl multichannel pipette.

Libraries were then normalized to allow equal library representation in pulled samples for sequencing. Library normalization buffer was prepared by mixing library normalization additives LNA1 (Illumina, San Diego, USA) with library normalization beads LNB1 (Illumina, San Diego, USA). Forty five microlitres of this normalization buffer was then added to each well of the PCR 96 well plate containing the PCR cleaned libraries. The plate was shaken at 1800rpm for 30 minutes then placed on a magnetic stand until the liquid was clear and a pellet was seen at the bottom of each well. Supernatant was discarded using a 20µl pipette and the beads were washed twice by adding 45µl of library normalization wash buffer NLW1 (Illumina, San Diego, USA) to each well and shaking for five minutes at 1800rpm. After shaking the plates were placed on a magnetic stand until the liquid was clear and the supernatant was discarded with a 20µl pipette. This was process was then repeated. Following washing 30µl of 0.1M sodium hydroxide was added to each well, to elute the DNA from the normalization beads, and the plate was shaken at 1800rpm for five minutes. Following shaking the plate was placed on a magnetic stand for two minutes to allow liquid to clear. Supernatant liquid was transferred from the midi plate to a PCR clean 96 well plate.

Prepared libraries were then pooled by aliquoting 5µl of liquid from each well of the 96 well plate into a 1.5ml sterile Eppendorf. The pooled library was then diluted to 6-8pM in hybridization buffer HT1 (Illumina, San Diego, USA) as described by the Illumina protocol. Samples were loaded into the sample well of a MiSeq reagent cartridge (Illumina, San Diego, USA) and loaded onto the MiSeq sequencer (Illumina, San Diego, USA) for 2x150bp paired end reads to obtain a minimum coverage of 20x for each

genome. A negative control of no DNA was included in all sample preparation and sequencing runs.

2.7 Analysis of WGS data

A bioinformatics pipeline was constructed to determine SNP occurring in the core genome from the raw Illumina data and produce maximum likelihood (ML) phylogenetic trees from the derived SNPs. The pipeline was set up to exclude SNPs occurring in highly repetitive regions of the genome, MGEs as well as SNPs that may have arisen by homologous recombination.

In order to undertake Bayesian analyses, SNP alignments produced by the pipeline were applied to a Bayesian analysis platform to generate maximum clade credibility (MCC) trees.

2.7.1 Development of Bioinformatics pipeline

2.7.1.1 Pipeline production and implementation

Published literature and public bioinformatics forums were searched to find the most suitable programmes to construct a WGS pipeline that would check raw Illumina sequenced data for quality, align data to a reference genome, call SNPs and create phylogenetic trees. Literature and bioinformatics forums were further investigated for specific programmes and previously published pipelines to determine appropriate sources and how best to implement them into a custom pipeline. Programmes were

selected on their performance, speed, accuracy and ability to produce desirable output files and data.

All 437 genomes were applied to a finalized pipeline consisting of twelve programmes (fastQC, trimmomatic ,BWA, samtools, qualimap, bcftools, VarScan, vcfutils, seqtk, bedtools, gubbins and RAxML) and a single script to create dendrograms from raw sequencing data (Table 11)(**Error! Reference source not found.**).

Default parameters were used for FastQC, Trimmomatic, BWA, qualimap which quality checked, trimmed raw reads then mapped reads to the EMRSA-15 reference genome HO 5096 0412 (Koser *et al.*, 2012b); a representative isolate of EMRSA-15 that was isolated from a fatal neonatal infection in Suffolk, UK in February 2005. This reference genome was chosen for all bioinformatics mapping analysis of EMRSA-15 isolates as it is the only fully annotated ST22 genome available on GenBank (http://www.ncbi.nlm.nih.gov/genbank/) and has been used extensively in previous literature (Holden *et al.*, 2013, Hsu *et al.*, 2015, Harris *et al.*, 2013, Harrison *et al.*, 2014, Nubel *et al.*, 2013, Paterson *et al.*, 2015, Price *et al.*, 2014).

Samtools was used to generate pileup data to identify Genomic variants. SNPs and indels were identified with VarScan using separate commands (pileup2snp and pileup2indel) with minimum ten read coverage, two minimum supporting reads and minimum base quality of 30 parameters for both commands.

Bcftools "view" command was used to generate consensus vcf files for each genome and was piped with vcfutils (vcf2fq command), seqtk (seq command) and bedtools (maskfasta command) to create consensus fastas with repeat and MGE regions removed for each of the 437 genomes. Removed regions from each genome are detailed in Table 12. These 437 consensus fastas were merged to a single multi fasta file and

applied to gubbins for recombination region determination using default parameters. Identified recombinant, repeat region and MGE SNPs were removed from the SNP VCF file which was converted to a multi-fasta using the script "vcf2phyloviz" (https://github.com/nickloman/misc-genomics-tools/tree/master/scripts).

Phylogenetic reconstruction was achieved by applying the multi-fasta file to RAxML using a generalized time reversible (GTR) model and gamma distribution for among site variation. One thousand bootstrap replicates were conducted and applied to dendrograms.

2.7.1.2 Mobile Genetic Element analysis with Nullarbor

All 210 sequenced genomes were applied to the pipeline "Nullarbor" (Github https://github.com/tseemann/nullarbor) to determine MGEs occurring in each of the genomes. Untrimmed Illumina fastq files for each genome were applied to the pipeline which used Prokka (Seemann, 2014) for genome assembly and annotation and Roary (Page *et al.*, 2015) to determine the pan genome.

Programme	Version	Function	Reason for choice	Reference
FastQC	0.10.1	Sequencing quality control and validation	Fast analysis of raw data. Visual output of quality of data.	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Trimmomatic	0.32	Nextera sequencing adaptor trimming and read quality control	Fast and accurate analysis for Illumina data. Pre-loaded files for removal of Nextera adaptors	Bolger *et al* 2014
BWA	0.7.5	Read mapping to reference	Fast and accurate analysis for Illumina data	Li and Durbin 2009
Samtools	0.1.19	File manipulation and variant analysis	Standard programme.	Li *et al* 2009
Qualimap	2.1	Mapping quality and genome coverage analysis	Accurate values for mapping. Visual output for coverage.	Garcia-alcalde *et al* 2012
Varscan	2.4.0	Variant detection	Accuracy and determining SNP thresholds	Koboldt *et al* 2012
Bcftools	0.1.19	File manipulation, variant detection and analysis	Versatility of programme. Can produce bcf files and consensus sequences.	Li 2011
Vcfutils	0.1.19	File manipulation	Ease of use. Part of bcftools package.	https://github.com/lh3/samtools/blob/master/bcftools/vcfutils.pl
Seqtk	1.0	File manipulation	Ease of use	https://github.com/lh3/seqtk
Vcftools	0.1.12b	Removal of SNPs from vcf files	Ease of use	http://vcftools.sourceforge.net/
Bedtools	2.18.2	File manipulation. Removal of repeat regions from analysis.	Ease of use	Quinlan and Hall 2010
Gubbins	1.0.2	Homologous recombination detection and removal	Fast and automated recombination detection. Visual output for regions of recombination.	Croucher *et al* 2015
RAxML	8.0	Maximum likelihood phylogenetic tree construction	Robust programme for tree construction. Can generate bootstrap values to be added to trees.	Stamitakis 2014

Table 11: Programmes implemented in the bioinformatics pipeline and their function.

Element/Region (gene)	Genomic coordinates
SCC*mec*	34163..51525
IS element	80765..83026
Repeat region *(Spa)*	98547..98768
Transposase	137322..138041
Transposase	141474..142274
Repeat region (Staphylocoagulase)	236662..237098
IstB-like ATP-binding protein and transposases	315945..318143
Repeat region	633963..634193
IS element transposase	818385..820574
IS element	935317..936294
Transposase	938088..938798
IS element	977393..979583
IstB-like ATP-binding protein	1104852..1107038
IS element	1252874..1254311
IstB-like ATP-binding protein	1283817..1288790
putative transposase, putative insertion element protein	1328804..1330543
Repeat region	1361018..1375548
phiSLT and phi PVL like phage	1520143..1566314
IS element	1567574..1569770
IS element	1852864..1853851
Duplication	861820..1863920
IS element	1874756..1876610
IS element	1895353..1896040
IS element	1890687..1892984
Repeat region	1894737..1895351
IS element	1902192..1903454
Repeat region	1936053..1936070
IS element	1936131..1936301
IS element	1943176..1944547
IS element	1944548..1945123
IS element	1945380..1945465
Repeat region	1990382..1990425
Tn552-like	2021753..2029411
IS element	2006700..2008180
Prophage (PhiSa3)	2042948..2087965
Phage terminase	2098645..2099998
Repeat region	2103706..2103983
Transposase	2136387..2138610

Table 12: Regions of the EMRSA-15 genome removed from SNP analyses.
Continued on page 78.

Element/Region (gene)	Genomic coordinates
Repeat region	2407740..2408448
IS element	2509256..2510560
Transposase	2568107..2570285
IS element	2585866..2586861
Intergenic	2631410..2633180
Transposase	2700680..2701410
IS element	2774816..2778254
IS element transposase	2825393..2826315

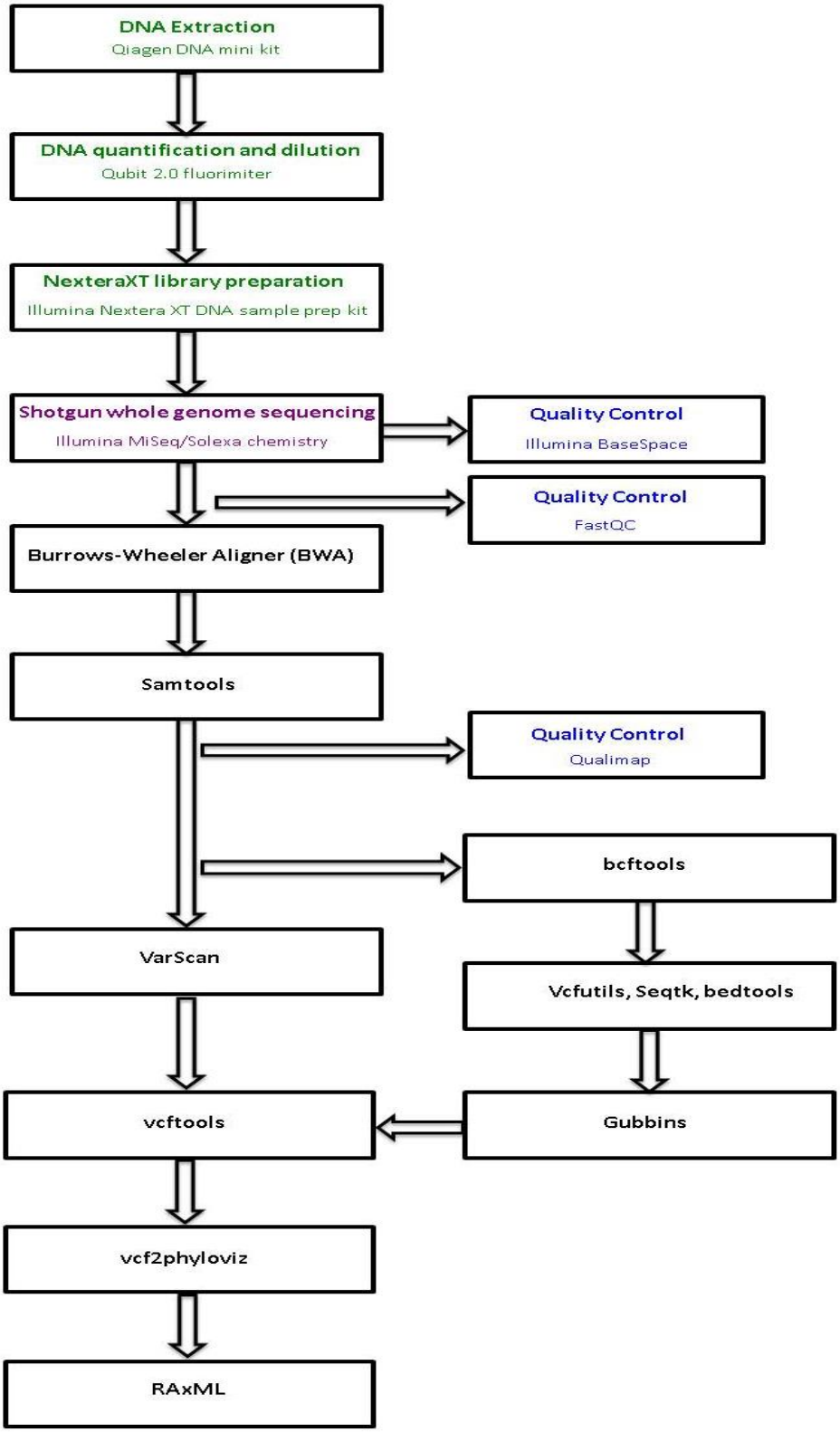

Figure 5: Pipeline of DNA extraction, sequencing and Bioinformatics processing. Bench work stages are coloured green, sequencing purple, quality control blue and bioinformatics black.

2.7.2 Bayesian Inference

The Bayesian analysis package BEAST (Bayesian evolutionary analysis sampling trees) 1.8.1 (Drummond *et al.,* 2012) was used to determine the spatiotemporal evolution of the SNP dataset produced by the pipeline. BEAST is a platform of six programmes (Table 13) that are implemented in a stepwise process to produce an annotated maximum clade credibility (MCC).

2.7.2.1 File generation and BEAST implementation

The ML tree produced with RAxML was uploaded to the programme Path-O-Gen 1.4 to determine the molecular clock rate of the dataset. Each isolate in the tree was tagged with a sampling date to allow the evolutionary rate to be calculated.

A graph of root to tip distances of each isolate against time and a dendrogram identifying isolates that do not adhere to the strict molecular clock rate of isolates was produced. Evolutionary rate values were automatically calculated by Path-o-Gen which are required for input in BEAST.

Stage/Step	Programme	Version	Function
1	Path-O-Gen	1.4	Testing clock rate of phylogenetic trees. Estimation of evolutionary rate.
2	BEAUTi	1.8.1	Preparation of XML files for BEAST input. Setting of evolutionary rate model, tree model and priors for BEAST implementation.
3	BEAST	1.8.1	Bayesian analysis of XML files
4	Tracer	1.6	Analysis of log files from BEAST
5	Tree Annotator	1.8.1	Construction of maximum clade credibility (MCC) trees
6	FigTree	1.4.0	Visualisation and annotation of MCC trees

Table 13: Programmes implemented within the BEAST software package.

An aligned multi fasta file of SNPs from all isolates was uploaded to the programme BEAUTi 1.8.1 in order to set parameters for the main BEAST programme and generate appropriate XML files for BEAST input. These included selecting of site substitution models, clock rate models, population models and priors that affect how the BEAST Bayesian inference is run.

In order to determine the best model for the dataset all combinations of strict, relaxed lognormal and relaxed exponential clock models with constant, exponential, expansion and Bayesian skyline population models were evaluated in triplicate runs. All combination pairs of clock and population models used a GTR model with gamma correction for among site variation site model and included the sampling dates of isolates in the analysis. Priors were set to default with the exception of the "ucld.mean" prior, which was set to a normal distribution using the rate value obtained from Path-o-gen as the initial and mean input values. XML files generated by BEUTI were applied to the BEAST programme and each replicate was run for 100,000,000 chains and sampled every 10000 generations.

2.7.2.2 Quality analysis and interpretation

BEAST runs were confirmed for convergence and checked for quality using Tracer 1.6. Convergence was confirmed by having effective sample size (ESS) values equal to or greater than 200 for key parameters including posterior, prior and likelihood statistics. Convergence was further confirmed by a "hairy caterpillar" trace.

The Tracer model comparison tool Bayes factor test was used to determine the best model to fit the ST22 population dataset using an aimc model and 1000 bootstrap replications.

Once the best model had been selected a burn in of 10 million chains was removed from each replicate run before all three triplicate runs for each clock/population model were combined with the LogCombiner V1.8.1 programme. The resulting combined trees file was used to create MCC trees with TreeAnnotator V1.8.1. MCC trees were viewed and edited with FigTree V1.4.0 to colour branches by sampling origin, show divergence dates and apply scaled axes to the trees.

2.7.3 MLST typing

All sequenced genomes were confirmed as ST22 by multi locus sequence typing (MLST) via submission of fastq files to the center for genomic epidemiology (CGE) server (https://cge.cbs.dtu.dk/services/MLST/).

2.8 VNTR and WGS comparison

VNTR and WGS methodologies were compared in two ways. Firstly the number of tandem repeats at VNTR loci L13 and L21, the two most variable loci in EMRSA-15, were applied at the tips of a BEAST constructed MCC tree using the Interactive Tree of Life (ITOL) website (http://itol.embl.de/).

Secondly differences in SNP and VNTR typing methods were compared. Each of the 210 WGS sequenced EMRSA-15 isolates were compared to each other isolate for the number of SNP differences, number of VNTR locus variants (LV) and summed tandem repeat difference (STRD). Multi-fasta files produced from the bioinformatics pipeline detailed in section 2.7.1 were applied to MEGA 5.2.2 to generate a pairwise SNP distance matrix of the 210 isolates which had also been VNTR typed. This matrix calculated the SNP distance of each of the 210 isolates to every other isolate in the multi fasta file. The matrix was converted to a list of "isolate x" versus "isolate y" SNP distances in an excel format output from MEGA. All versus all pairwise LVs and STRDs were calculated using Microsoft Excel. LV was defined as the number of loci each isolate varied at (relative to every other isolate). STRD for each isolate was calculated by the sum of the absolute difference in tandem repeat copy number at each locus. Pairwise values were then calculated as the difference in STRD between isolates. Data for LV, STRD and SNP differences were concatenated using Microsoft Excel to give a table of the SNP distance, LV and STRD of every isolate to each other. Data was applied to STATA 14.2 to generate SNP versus LV and SNP versus STRD graphs with increasing data point size representing increasing number of isolates.

2.9 Antimicrobial sensitivity testing

Eleven isolates were tested for sensitivity to ciprofloxacin, levofloxacin and ofloxacin using Etests (Biomerieux, Marcy-l'Etoile, France). Isolates were selected from their positioning in the EMRSA-15 global phylogeny described in section 3.2. These consisted

of three randomly selected BHM clade isolates, three of the isolates basal to the BHM clade, three randomly selected isolates out with the BHM clade and single isolates from Guernsey and Hong Kong.

MRSA from overnight cultures grown on Muller Hinton agar at 37°C were suspended in 5ml of ISO broth to match a 0.5 McFarland standard followed by suspensions being diluted in a ratio of 1:10 with sterile water. Adjusted suspensions were used to create a bacterial lawn on ISO sensitest agar and sterile forceps were used to place an Etest strip (range 0.002 to 32 µl per ml for all three antibiotics) onto the dried lawn. One Etest strip per bacterial lawn/agar plate was applied. *S. aureus* control strains ATCC 29213 and NCTC 1066 were used as sensitive and resistant controls respectively. All plates were incubated at 37°C for 24 hours before the diameter of zones of inhibition were measured against the Etest strip markings. MIC values above one mg/L were considered resistant and below one mg/L were considered sensitive (European Comittee on Antimicrobial Susceptibility Testing (EUCAST), 2016). *S. aureus* strains ATCC 29213 and NCTC 1066 were used as sensitive and resistant controls respectively.

2.10 Metadata acquisition and analysis

2.10.1 Bacteraeamia and occupied bed days (OBD) data

Voluntary and mandatory annual MRSA bacteraemia data for HEFT between 1993 and 2015 was retrieved from Public Health England (PHE) (https://www.gov.uk/government/statistics/mrsa-bacteraemia-annual-data).

Occupied bed days for HEFT between 2001 and 2015 were retrieved from the HEFT hospital record system.

2.10.2 Antibiotic consumption data

Annual in-patient antibiotic prescribing data for fluoroquinolones, macrolides and third generation cephalosporins (3GCs) between the years 1990 and 2015 were obtained from the pharmacy record system at HEFT. Quantities of each drug were converted to number of defined daily doses (DDD) as defined by the World Health Organization (WHO) collaborating Centre for Drug Statistics Methodology (http://www.whocc.no).

2.11 Desiccation assay

Isolates of *S. aureus* were tested for their tolerance to desiccation under constant relative humidity (RH) and temperature in a sealed environment. Glass coverslips were inoculated with bacteria and enumerated at set time points of 0, 12 hours and 1, 2, 3, 5, 7, 10, 14, 21, 28 and 35 days post inoculation and enumeration counts used to determine the survival of each *S. aureus* strain over time.

2.11.1 Bacterial strain selection

Nine isolates were selected for inclusion in the desiccation assay based on phylogenetic results from chapter 3.2 along with the NCTC EMRSA-15 and MSSA-476 control strains as detailed Table 14. Isolates were selected to include clinical and screening isolates and

comprised three isolates from a localized, dominant Birmingham strain, three isolates that were sporadically distributed with other UK and two single isolates from Hong Kong and Kuwait that were chosen to determine the desiccation tolerance of EMRSA-15 from non-UK sampling regions. The EMRSA-15 NCTC control strain (NCTC 13142) was included to act as an EMRSA-15 control and the NCTC MSSA-476 strain (NCTC 13297) was included to act as an MRSA and MSSA comparator.

2.11.2 Desiccation protocol validation

2.11.2.1 Sterilization of glass coverslips

Five glass coverslips were placed in an open petri dish in a PCR clean hood and irradiated with UV light at 450nm for 15 minutes and another five glass coverslips were immersed in 98-99% ethanol for two minutes then left to dry in a sterile petri dish in a PCR clean hood. The five UV irradiated coverslips were placed into individual brain heart infusion (BHI) (Oxoid, Basingstoke, UK) broth and vortexed for 30 seconds. 100µl of broth was inoculated onto a Colombia blood agar (CBA) (Oxoid, Basingstoke, UK) plates and both BHI broth and CBA were incubated at 37^{o}C for 24 hours. This was repeated for the five ethanol washed coverslips and again for five coverslips that had undergone no sterilization procedure. A negative control of BHI broth with no coverslips and a positive control of the NCTC EMRSA-15 strain (NCTC 13142) inoculated BHI were included. After incubation plates and broths were inspected manually for growth and any broths showing no growth were incubated again for 24 hours at 37^{o}C and inspected after incubation.

2.11.2.2 Maintenance of relative humidity (RH) and temperature

The maintenance of a stable RH in a sealed environment was verified. A digital thermo-hygrometer (Fischer Scientific, Hampton, USA) was placed in a large sterilized plastic box (height 23cm, width 44cm, depth 71cm) that was sealed and humidity and temperature readings were taken every hour for 8 hours and then again at 24 hours to establish the baseline temperature and RH. To establish if a saturated salt solution could maintain a relative humidity, the thermo-hygrometer was sealed in the box with a saturated solution of calcium chloride (Sigma-Aldrich, St Louis, USA) (5g of calcium chloride in 5mls of water) in a 100ml beaker and temperature and humidity readings were taken every hour for 8 hours and then again at 24 hours.

2.11.3 Inoculation of glass coverslips with bacteria

A single colony pick from an overnight growth of each strain that was grown on ISO-Sensitest agar (Oxoid, Basingstoke, UK) at 37^{O}C was mixed with 10ml BHI broth (Oxoid, Basingstoke, UK) and incubated at 37^{O}C overnight. Overnight broth cultures were quantified using the Nanodrop (Thermo Scientific, Waltham, USA) for optical density (OD) at a wavelength of 600 nm and then diluted to an OD of 0.1. A 1ml aliquot of each diluted culture was pipetted into a sterile 1.5ml Eppendorf tube (Eppendorf, Hamburg, Germany) and centrifuged at 13,000rpm for five minutes to pellet cells. Supernatant was collected and discarded then replaced with 1ml of either water or 3% bovine serum albumin (BSA) and re-suspended by vortexing for 30 seconds. BSA was used as comparator menstrua to water as it acts as a protective media that has been used

previously to represent dirty biological conditions at 3% w/v solution (Kawamura-Sato *et al.*, 2008, Tuladhar *et al.*, 2012).

Three UV irradiated glass coverslips per time point were inoculated with 20µl of either water or BSA suspended bacteria, giving a total of 36 inoculated coverslips per strain that were placed inside a sterile petri dish which was in turn placed in the desiccation chamber.

A saturated salt solution of 5g of calcium chloride in 5mls of water was placed in the desiccation box and left overnight to establish a controlled RH. Petri dishes containing the inoculated coverslips, one petri dish per *S. auerus* strain, were placed inside the box and the calcium chloride solution was replaced with fresh solution. A digital thermo-hygrometer (Fischer Scientific, Hampton, USA) was placed in the box to monitor temperature and RH then the box was sealed. The box was opened at each designated time point to remove coverslips for enumeration. Saturated salt solutions were replaced each time the box was opened.

ISOLATE	Sampling Location	Ciprofloxacin sensitivity	Clinical or screening	BHM clade
CT124	Birmingham	Sensitive	Unknown	no
HK_N5	Hong Kong	Resistant	Screen	no
KW64±	Kuwait	Sensitive	Unknown	no
A120	Birmingham	Resistant	Screen	yes
ECMM195	Birmingham	Resistant	Screen	yes
CT147	Birmingham	Resistant	Clinical	yes
A901	Birmingham	Resistant	Screen	no
EMM513	Birmingham	Resistant	Screen	no
CT300	Birmingham	Resistant	Clinical	no
EMRSA-15 control	-	Resistant	-	no
MSSA-476 control	-	Sensitive	-	no

Table 14: Isolates included in the desiccation assay.

2.11.4 Enumeration of viable bacterial counts

At each time point three inoculated glass coverslips from each strain were removed from the desiccation box and were each eluted in separate 2ml volumes of phosphate buffered saline (PBS) in a 15ml polystyrene universal (Thermo Fischer Scientific, Madrid, Spain) and vortexed for 15 seconds.

Eluents were then appropriately diluted in PBS to allow suitable growth on agar plates for enumeration and applied to a spiral plater (Don Whitley Scientific, Shipley, UK) which deposited bacterial solutions onto ISO sensitest agar using a 50µl log deposition in a concentric spiral from the centre of a round agar dish outwards. The methodology was altered over time to accommodate the reducing number of viable bacteria being enumerated. Day 0 to day 14 water suspension viable counts were determined by washing each glass coverslip in 2ml (PBS) and application to the spiral plater. This was reduced to 1ml eluent with flood plate methodology from day 21 onwards. Flood plates had 100µl volumes of eluent deposited directly onto ISO sensitest agar plates and spread with a sterile plastic spreader. Due to the slower reduction in bacteria surviving in BSA this methodology change occurred at day 35 in the BSA suspension assay with time points 0 to 28 being applied to the spiral plater.

Three replicate plates were spiral plated or flood plated for each coverslip to give a total of nine replicates per time point. Agar plates were incubated at 37^{o}C for 24 hours before being enumerated using a spiral plate counting grid or total colony counts on flood plates. Colony counts for each replicate were recorded and then used to calculate the

number of bacteria surviving on each glass coverslip by averaging the values for each of the nine replicates.

2.11.5 Analysis

Standard deviation and standard error were calculated for each time point in water and BSA using the respective functions in Microsoft Excel.

Percentage survival values for each time point were generated by dividing the average colony count of nine replicates at each time point by the average colony count for time point 0. Area under the curve values were calculated using Microsoft Excel using the trapezoidal equation y = f(x).

P-values were calculated for each time point by comparing the nine colony count values of isolates suspended in water to the nine colony counts of isolates suspended in BSA using a two-tailed Student's T test.

Exponential linear regression graphs were constructed with Microsoft excel and had error bars for each time point applied.

Whole genome sequencing data for each isolate was scrutinized for SNPs or indels occurring in the core genome that may be linked to desiccation tolerance. A vcf file was generated for the eleven genomes investigated using the bioinformatics pipeline described in section 2.7.1.1. Fourteen key genes with known stress response or desiccation tolerance activity (Table 15) were scrutinized. Validation of SNP position and coverage was accomplished by uploading sorted bam files to Artemis comparison tool V15.0 and comparison with the EMRSA-15 reference genome annotated GenBank file.

Gene	Genomic location in EMRSA-15 reference genome	
	Start	Finish
sodM	124261	124860
ahpC	397470	398039
SigH	548176	548745
spx	937591	937935
KatE	1309029	1310483
sodA	1627085	1627684
clpX	1738081	1739343
perR	1953517	1953963
agr	2109592	2109861
SigB	2148019	2148789
rsbU	2149691	2150692
SarR	2366747	2367094
crtN	2661629	2663137
crtM	2663149	2664012

Table 15: Genes linked to desiccation tolerance that were investigated for mutation at each of the eleven isolate genomes investigated.

3 Results

3.1 Whole genome sequencing and bioinformatics pipeline validation

A total of 317 MRSA isolates sampled across five hospital trusts from the UK, Guernsey and Hong Kong were successfully cultured and VNTR typed (Table 8). Two hundred and ten of these isolates were determined to be EMRSA-15 with a VNTR profile highly related to the EMRSA-15 control strain, with one, zero and two repeats at VNTR loci L1, L13 and L7 respectively (Appendix [Table 25]).

In addition to these sequenced isolates, 227 previously sequenced and published genomes from isolates that had been sampled across the UK and around the world (Table 10) were downloaded in fastq format and included in all bioinformatics analyses (Holden *et al.*, 2013, Reuter *et al.*, 2016).

3.1.1 Whole genome sequencing results

The two hundred and ten isolates were successfully sequenced to an average coverage of 60.10x (range 15.0.x to 428.7x). All isolate DNA sequences were confirmed to have sufficient quality for bioinformatics analysis, with quality score of 28 or above using FastQC.

All sequenced genomes were confirmed as ST22 by being applied to the CGE MLST typing server and produced the allelic profile of 7 6 1 5 8 8 6 for the *arcC*, *aroE*, *glpF*, *gmk*, *pta*, *tpi*, and *yqiL* genes respectively.

3.1.2 Bioinformatics pipeline outputs

The raw fastq files of the 210 sequenced isolates along with the 227 downloaded fastq format genomes were applied to the bioinformatics pipeline to give a total of 437 analysed genomes. Automation of the pipeline was achieved for over 75% of command inputs, reducing the chance of error for manual command input and reducing user interface time. Twenty five percent of commands could not be automated due to programmes not being compatible for piping (linking together).

Qualimap demonstrated complete read mapping across the reference strain for each genome and was used to confirm read coverage. After SNPs arising in MGEs and repetitive regions were removed a total of 13293 SNPs were determined from the pileup file of the 437 genomes in a vcf.

Consensus sequences for each genome were successfully generated and combined into a single multi-fasta and applied to gubbins to determine regions of homologous recombination across the dataset. Gubbins converged on a recombination removed phylogeny after five iterations and identified 50 recombination blocks of between four and 18 SNPs that gave a total of 419 recombinant SNPs in the dataset (3.15% of all SNPs). This would indicate homologous recombination across the 437 isolates was low.

These recombinant SNPs were removed from the MGE and repetitive region clean vcf (containing 13293 SNPs) to give a total of 12874 core SNPs across 437 genomes.

The 12874 SNPs in the vcf file were concatenated to a multi-fasta file and applied to RAxML which successfully generated a ML tree with 500 bootstrap iteration values applied to it (Figure 6). Bootstrap values ranged from 0 to 100, with over 80% of values ranging from 50 to 100 (mode 100). These high ranges of bootstrap values are indicative of a robust phylogenetic reconstruction, meaning the tree produced by the pipeline would be appropriate for application to path-o-gen for determination of molecular clock rate (Section 3.1.3.1).

The pipeline developed was compared to five pipelines that have been recently used to analyse MRSA WGS (Holden *et al.*, 2013, Reuter *et al.*, 2016, Hsu *et al.*, 2015, Aanensen *et al.*, 2016, Harris *et al.*, 2010). Different programmes and methodologies were used for all key steps with the exception of phylogenetic tree reconstruction (Table 16). The application of Bayesian phylogenetics to generate time linked phylogenetic trees was also only applied in two other studies (Holden *et al.*, 2010, Hsu *et al.*, 2015) investigating ST22 epidemiology in the past five years but also used different models for tree construction compared to this study.

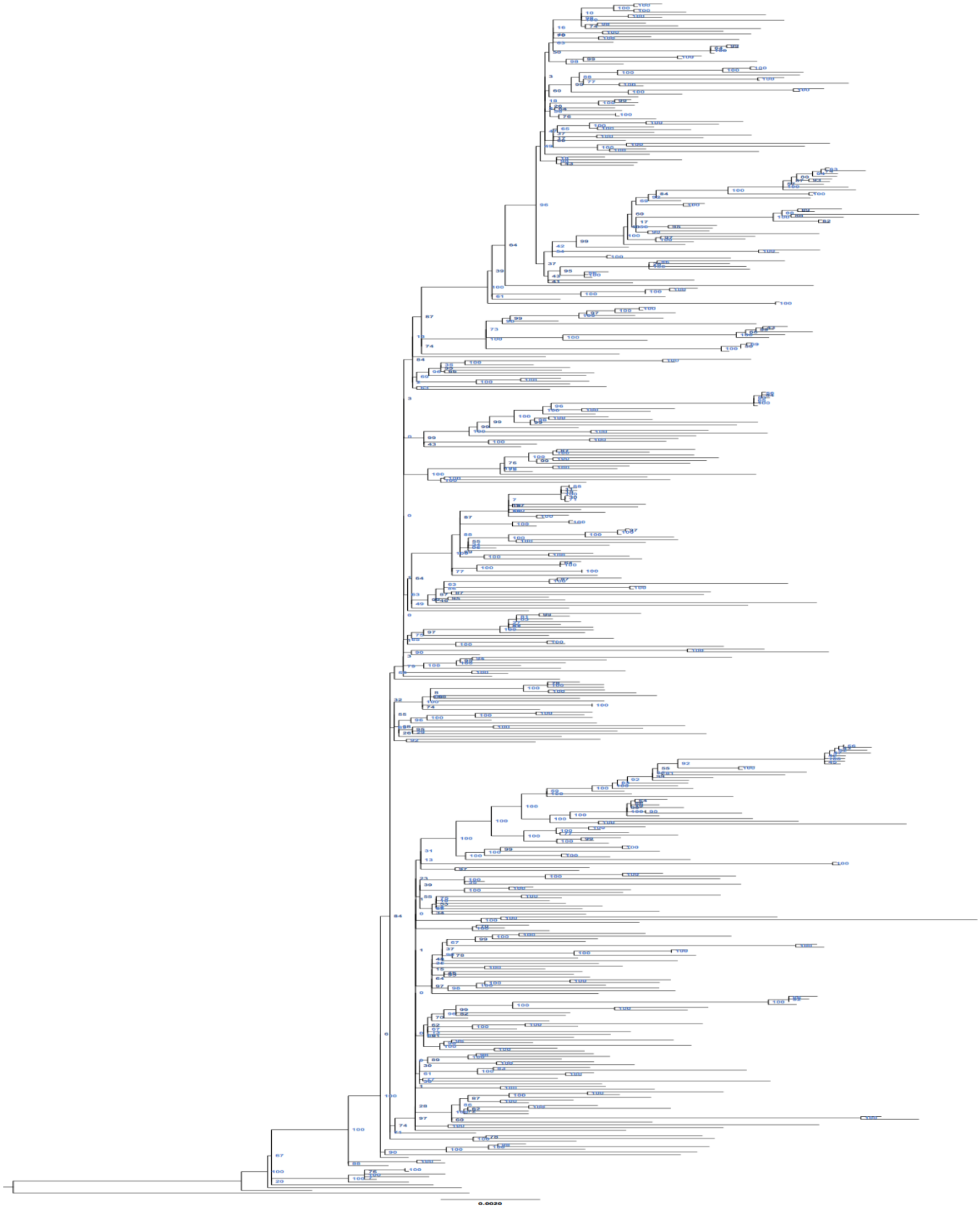

Figure 6: Maximum likelihood (ML) phylogenetic tree of the 437 genomes included in the study. Bootstrap values are displayed at each node of the tree in blue.

Process		This Thesis	Reuter *et al* 2016*	Aanensen *et al* 2016*	Hsu *et al* 2015*	Holden *et al* 2013	Harris *et al* 2010
Isolates/Genomes	**Number**	437	783	41	87	193	43
	Locations	Global	UK	Europe	Singapore	Global	Global
Sequencing	**Platform**	Illumina MiSeq	Illumina HiSeq	Illumina Genome Analyser GAII, Illumina HiSeq	Illumina Genome Analyser GAII, Illumina HiSeq	Illumina Genome Analyser GAII, Illumina HiSeq	Illumina Genome Analyser GAII
	Reads (bp)	2x 150bp	100bp	54bp, 75bp	Not stated.	54bp, 75bp	54bp
Mapping	**Programme**	BWA	SSAHA	SSAHA	SMALT	SSAHA	SSAHA
	Reference genome	HO 5096 0412	HO 5096 0412	HO 5096 0412	HO 5096 0412	HO 5096 0412	TW20
Variant calling	**Programme**	VarScan, bcftools	SSAHA	SSAHA	Samtools	SSAHA	SSAHA
	SNP threshold	10x coverage, min 2 supp. reads.	Quality score >30	Quality score >30	Read depth 5, SNP score >60	Quality score >30	Quality score >30
Recombination removal	-	Gubbins	N/A	N/A	Gubbins	Manually removed.	N/A
Phylogenetics	**Programme**	RAxML	RAxML	FastTree, RAxML	RAxML	RAxML	RAxML
	Method	GTR	GTR	Not stated.	GTR	GTR	GTR
Bayesian Inference	-	BEAST	N/A	N/A	BEAST	BEAST	N/A
	Model	Relaxed clock constant population	-	-	Relaxed clock skyline population	Relaxed clock	-

Table 16: Comparison of programmes and parameters used in bioinformatics pipelines applied to MRSA WGS studies. Harris *et al* 2010 did not analyse EMRSA-15 genomes but is included as Holden, Aanensen and Reuter papers use SNP calling parameters outlined in this study. *Publications included non ST22 isolates. Information displayed is relevant to ST22 isolate analysis.

3.1.3 Bayesian analysis results

3.1.3.1 Path-o-gen results

The phylogenetic tree produced in section 3.1.2 was applied to the BEAST platform programme Path-o-gen to determine the molecular clock rate of the dataset**, Figure 7** shows each isolate in the tree's root to tip divergence against time. A scatter plot of each of the 437 isolates distance from the strict molecular clock (Appendix [Figure 30]) was used to determine cut offs for isolates displaying fast or slow evolutionary rate relative to a strict clock. Isolates falling out with a normal distribution (beyond residual values of plus or minus 2.0E-3) from the strict molecular clock line were considered outliers. These cut off points were applied to Figure 7 as dotted lines above and below the solid strict molecular clock line. Figure 7 shows eighteen isolates above the upper dotted line (displaying a fast rate of evolution) and fourteen isolates below the lower dotted line (displaying a slower rate of evolution) compared to the strict molecular clock line. None of these outlier isolates showed poor mapping or irregular sequences and did not cluster together but were randomly distributed throughout the tree as displayed in figure 8. The remaining 411 isolates were within acceptable distance of the clock rate line to be considered to have a stable evolutionary rate. Due to a low number of outlier isolates (n= 26, 5.9% of isolates) the rate value of 3.11628E-4 produced by Path-o-gen was considered acceptable as a rate input value for BEAST and that a relaxed molecular clock would be appropriate for BEAST analyses.

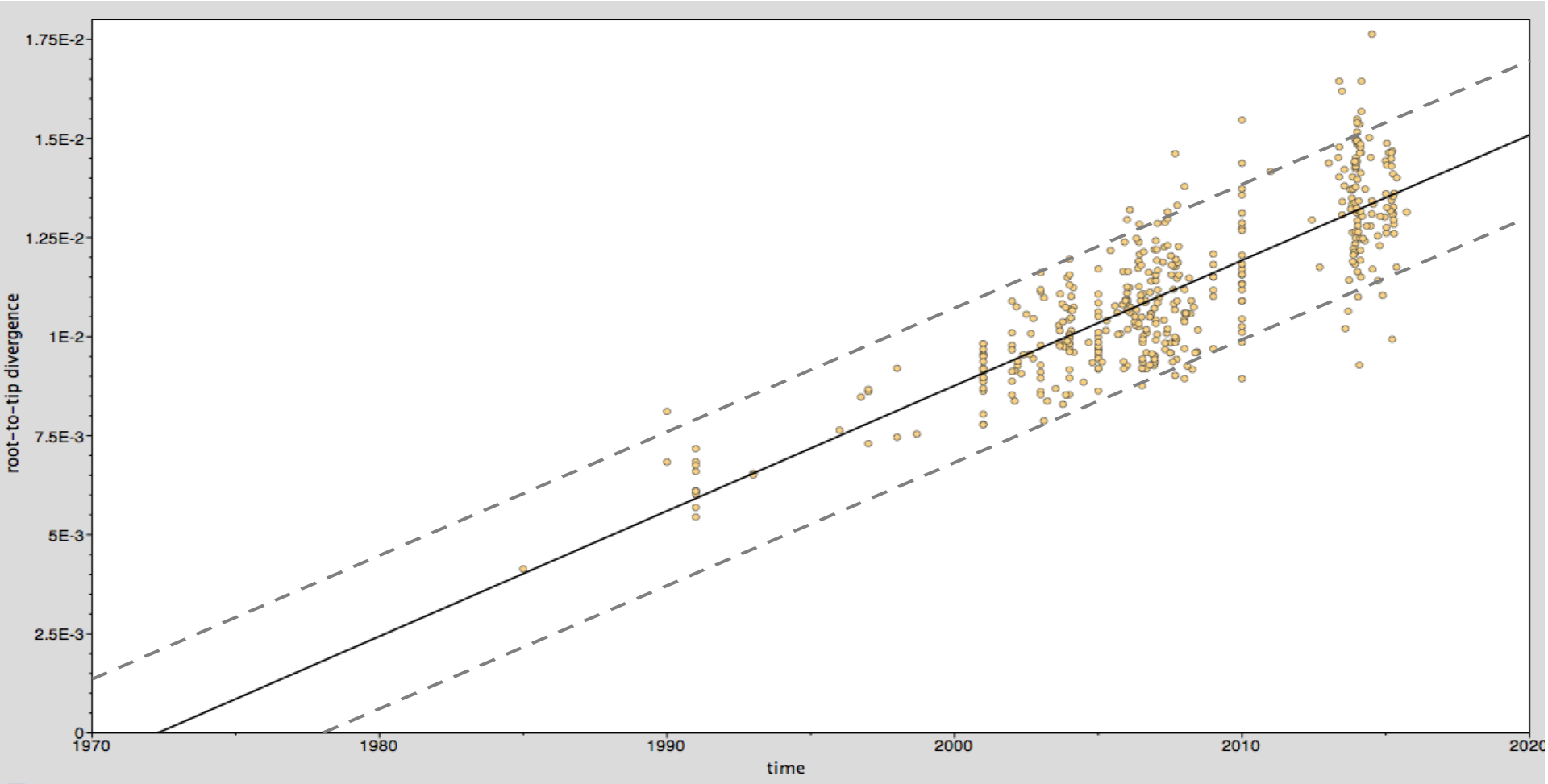

Figure 7: Plot of root to tip divergence values for each isolates in the tree against time. The solid black line represents a strict molecular clock rate. Isolates above the line have a faster divergence rate that the fixed molecular clock and isolates below the line have a slower divergence rate. Dotted lines indicate the normal distribution cut off points, with isolates between the dotted lines being within the normal distribution.

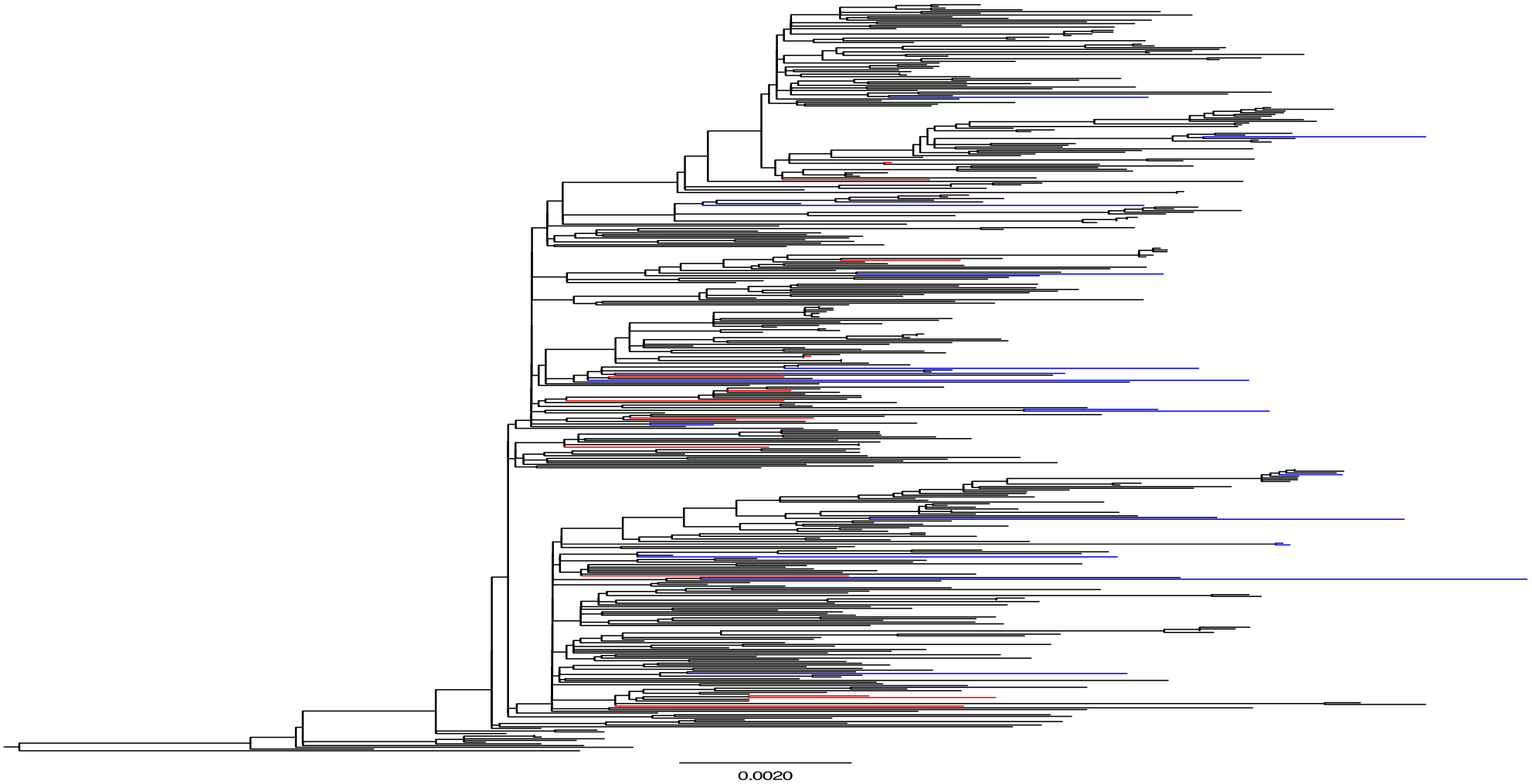

Figure 8: ML phylogenetic tree identifying outlier isolates with fast or slow evolutionary rate. Branches in black fit to a strict clock model. Branches coloured red have evolved slower than the fixed clock rate and branches colored blue have evolved faster than the fixed clock rate.

3.1.3.2 Bayesian Inference

An aligned fasta file of 12874 SNP sites that excluded SNPs in MGEs and repetitive regions or those arising by homologous recombination was applied to BEAST to generate .tree and .log files for MCC tree generation and quality analysis.

Run combinations including exponential clock models or Bayesian skyline population models failed to converge and were discarded, therefore leaving runs consisting of strict and relaxed lognormal clock models with constant, exponential or expansion population models. All model combinations were confirmed to have converged by having effective sampling size (ESS) values greater than 200 and demonstrating "hairy caterpillar" traces (Figure 9).

Bayes factor test was used to determine the model that best fit the dataset (Table 17). The model combination with the best fit was a relaxed clock model with a constant population model due to it having the lowest aimc value, and was subsequently chosen for all further analyses.

.tree files for the three independent runs of the lognormal relaxed clock with a constant size population model were successfully combined and applied to the Tree Annotator 1.8.1 programme to produce an MCC tree for phylogenetic analysis.

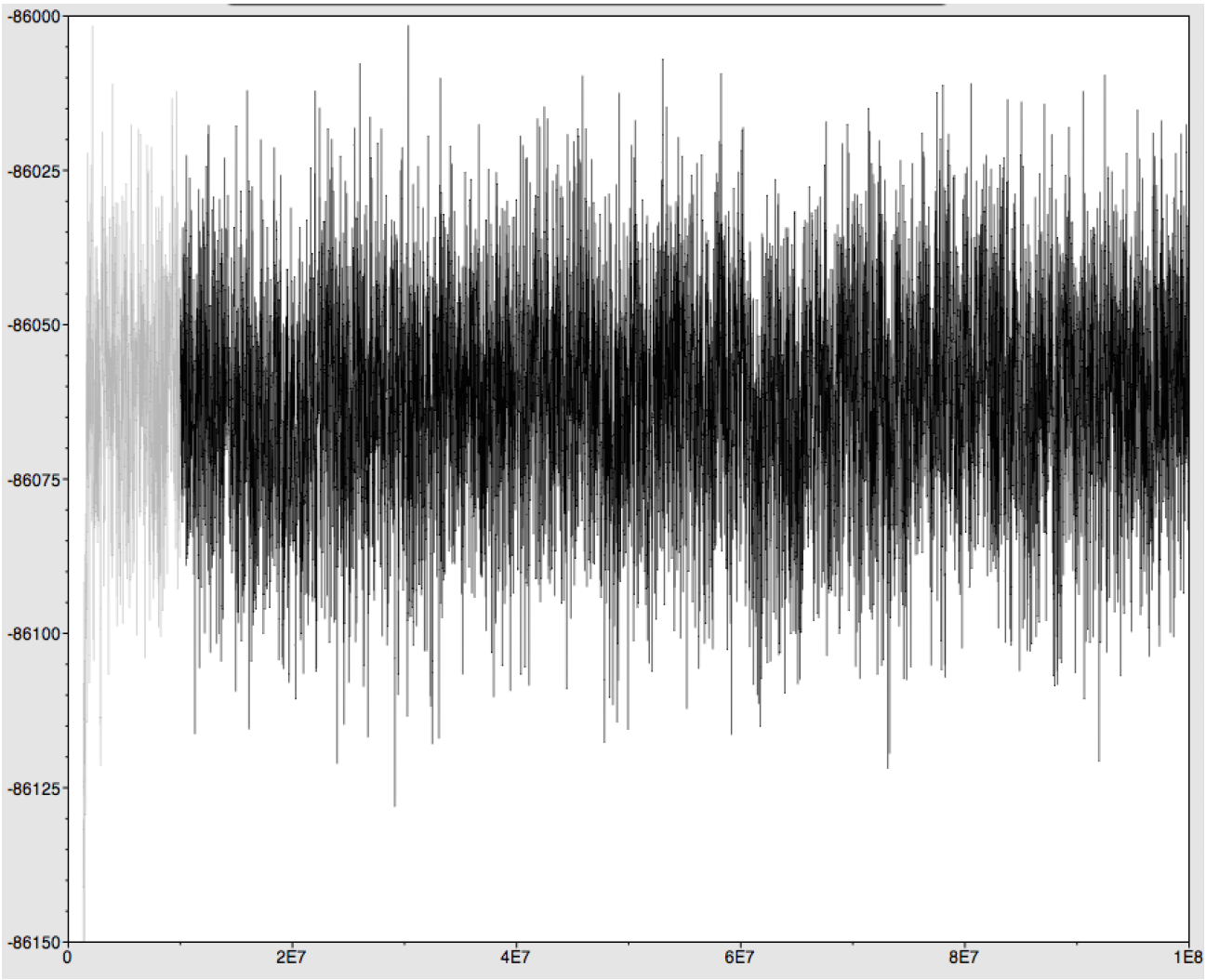

Figure 9: Example "hairy caterpillar" trace from one of the triplicate relaxed clock with constant population size models. White shaded area at the beginning of trace shows a 10% burn in at the beginning of the MCMC.

Run	AIMC	Relaxed constant	Relaxed expansion	Relaxed exponential	Strict constant	Strict expansion	Strict exponential
Relaxed constant	172204.523	-	-8.755	-8.252	495.971	475.469	469.602
Relaxed expansion	172213.278	8.755	-	0.503	504.726	484.224	478.357
Relaxed exponential	172205.026	8.252	-0.503	-	504.223	483.721	477.854
Strict constant	172709.249	-495.971	-504.726	-504.223	-	-20.502	-26.369
Strict expansion	172688.748	-475.469	-484.224	-483.721	20.502	-	-5.868
Strict exponential	172682.88	-469.602	-478.357	-477.854	26.369	5.868	-

Table 17: Table of Bayes factor test results comparing each convergent BEAST run to each other. Lower AIMC values indicate a better fit of the model to the dataset. Positive values indicate better relative fit of the rows model compared to the columns model.

3.1.3.2.1 Birmingham isolate MCC tree generation

To examine the phylogeny of EMRSA-15 in Birmingham in more detail, an additional MCC tree consisting of only isolates sampled from Birmingham (n = 181) was generated using BEAST. A fasta file of 5082 SNP sites was applied to BEAST and as with the BEAST run of the international collection in section 3.1.3.2 run combinations using exponential clock models or Bayesian skyline population models failed to converge. A lognormal relaxed clock with a constant size population model was also the best fit model using Bayes factor test.

3.2 The phylogeny of EMRSA-15

3.2.1 Evolutionary rate of EMRSA-15

The mean evolutionary rate for the core genome was determined to be 1.38 x 10^{-6} (95% Bayesian credibility intervals 1. 29x10^{-6} to 1.49 x10^{-6}) substitutions per nucleotide site per year.

3.2.2 The phylogeny of EMRSA-15

An MCC phylogenetic tree of 437 EMRSA-15 isolates was successfully generated by BEAST (section 3.1.3) and is displayed in Figure 10.

3.2.2.1 Evolution of ciprofloxacin resistance

A distinct phylogeny in respect to ciprofloxacin susceptibility was evident. The single ciprofloxacin sensitive historical CHB isolate clustered with the nine other ciprofloxacin sensitive genomes (from Holden *et al* 2013) at the base of the MCC tree (dotted box in Figure 10) and were separate from the 427 ciprofloxacin resistant genomes. This is highly consistent with the previous study that the ciprofloxacin sensitive genomes were sourced from (Holden *et al.*, 2013) despite using a different bioinformatics pipeline to determine SNPs in the core genome and a different Bayesian analysis model for MCC tree generation (Table 14).

The historical CHB isolate from c1985 (isolate CT124 [Table 25]) was the second most basal isolate on the tree and sat on a single branch that emerged in 1975 (95% Bayesian credibility intervals 1971.12 to 1978.9), two years before the larger cluster of

eight UK sampled ciprofloxacin sensitive isolate branches (diverged in 1977 [95% Bayesian credibility intervals 1974.47 to1981.57]) and nine years before the evolution of ciprofloxacin resistance. The CHB isolate was 43 SNPs from the nearest common ciprofloxacin sensitive ancestor node and 52 SNPs from the first reported EMRSA-15 genomes that were sampled from Birmingham six years after the historical CHB isolate.

Both the CHB historical isolate and the isolate directly basal to it lacked the deletion in the *ureC* gene that causes non-production of urease. This was determined indel identification with VarScan (position 2362301 G to GT indel) and manual inspection with Artemis.

All of the 427 ciprofloxacin resistant isolates clustered to a single large clade that emerged from the ciprofloxacin sensitive cluster on a single branch in approximately 1984 (95% Bayesian credibility intervals 1979.14 to 1985.14). The three isolates sampled from East Birmingham, where HEFT is located, between 1991 to 1992 (Richardson and Reith, 1993) were most basal isolates within the large ciprofloxacin resistant clade (Figure 10). This is again consistent with data from Holden and colleagues (Holden *et al.*, 2013). Considerable diversity was evident within the ciprofloxacin resistant clade. Four hundred and twenty isolates were distributed across two branches that that emerged in 1985 (95% Bayesian credibility intervals 1983.9 to 1987.4) and 1986 (95% Bayesian credibility intervals 1985.6 to 1988.4) respectively.

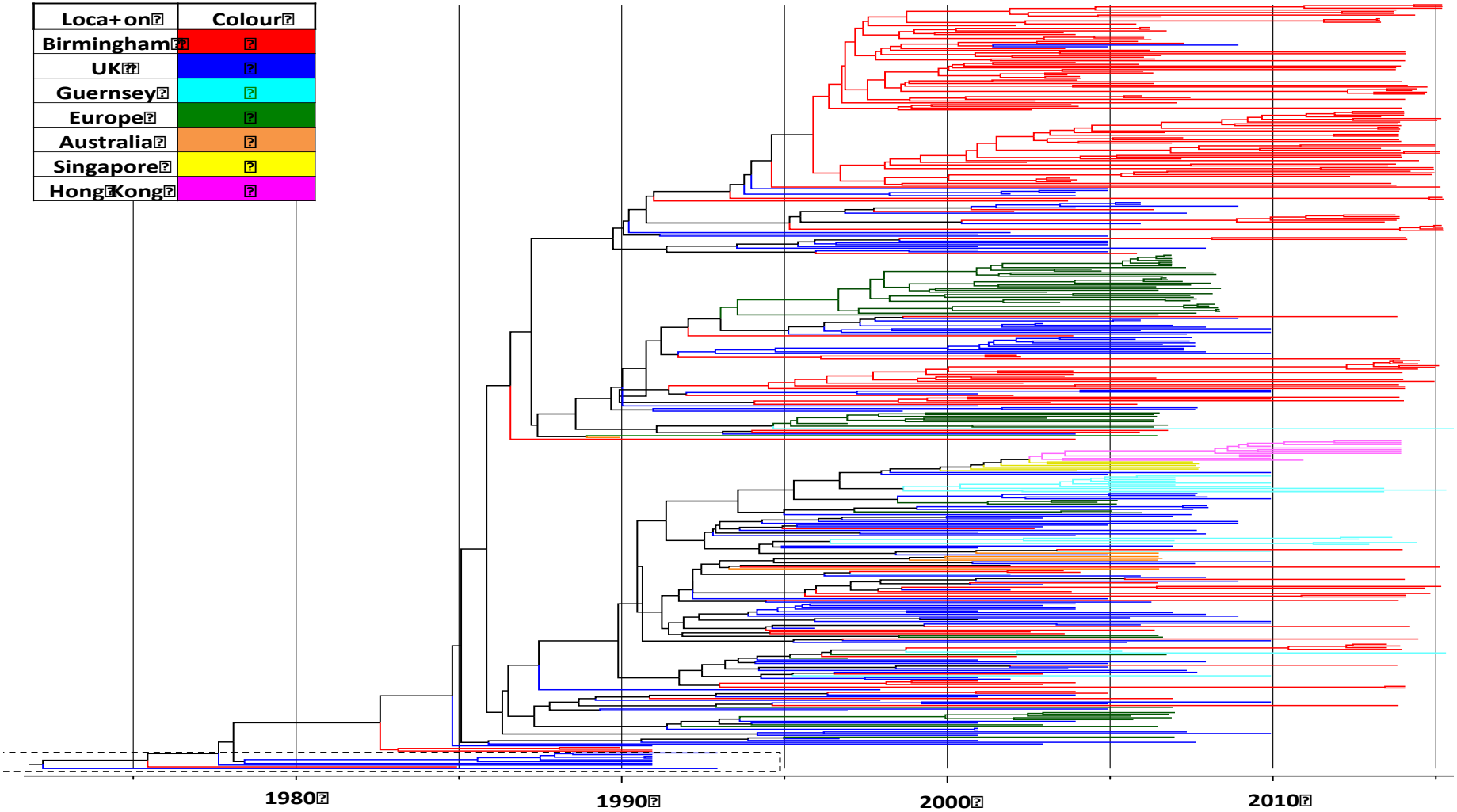

Figure 10: Maximum clade credibility (MCC) tree of the global EMRSA-15 population sampled between 1985 and 2015 coloured by sampling location. Branch tips are constrained by sampling time and time scale is displayed along the bottom (x axis). Colour legend at top left of figure. Ciprofloxacin sensitive isolates are highlighted in a dotted box at the base of the tree.

3.2.3 Global EMRSA-15 phylogeny

Substantial clustering of genomes based on country of sampling was evident. Non-UK genomes resolved to similar positions on the MCC tree as detailed in the analysis by Holden *et al* (Holden *et al.*, 2013) despite the inclusion of additional UK genomes. Clusters of German, Portuguese, Czech republic and Singaporean genomes were evident with each of these international isolate clusters (with the exception of Hong Kong) having isolates from the UK directly basal to their clusters (Figure 14) (Table 18). Of note, a single Guernsey sample was directly basal to the cluster of nine Portuguese genomes. Fourteen of the isolates from Hong Kong clustered together and arose from the Singaporean cluster (section 3.2.3.1). This was the only cluster of international isolates that did not have UK isolates located directly basal.

Region specific clustering of isolates to specific branches of the MCC tree was also evident with the Birmingham and other UK source genomes. A large cluster of Birmingham genomes was evident at the distal end of the tree (section 3.2.3.3) with additional smaller clusters displayed throughout the tree (section 3.2.4).

Twelve clusters of UK genomes, excluding Birmingham sampled isolates, comprising between three and 11 genomes grouping to single branches of the MCC tree were evident. Four of these branches comprised solely of isolates from the West Midlands (Reuter *et al* 2016) each of which had sequenced Birmingham genomes directly basal to each cluster. Two branches of mixed West Midlands and Birmingham isolates were evident, while the remaining eight branches of mixed UK locations comprised of Holden *et al* 2013 and Reuter *et al* 2016 genomes only.

3.2.3.1 Hong Kong

All 14 of the Hong Kong isolates clustered to a single branch of the MCC tree (Figure 10 and Figure 11) and showed evidence of hospital specific clustering. Eight temporally and geographically linked nasal screening isolates from neonatal wards at two hospitals (Appendix (Table 25)) clustered closely together with an average pairwise distance of 2 SNPs (range 0 to 4) suggesting these are highly related to each other. Five blood culture isolates from four hospitals were located on the same branch but were more distantly related both to each other and the linked screening isolates and had an average SNP distance of 36 SNPs (range 32 to 37). The blood culture isolate closest to the node containing all of the linked nasal screening isolates was distinguishable by 22 SNPs suggesting the blood culture isolates share a common ancestor but are more distantly related to the screening cluster. Overall, the average pairwise distance of all the 14 Hong Kong isolates to each other was 22 SNPs.

Country	Number of isolates in cluster	Location of basal UK isolates (Collection)	SNP distance to nearest common ancestor
Germany	35	Norfolk (Holden *et al* 2013)	24
Portugal	9	Guernsey	36
Czech republic	5	Shrewsbury, West Midlands (Reuter *et al* 2016)	36
Singapore	6	UK South East (Reuter *et al* 2016)	25

Table 18: Clusters of internationally sourced isolates with the UK isolates basal to each.

The branch containing all of the Hong Kong isolates emerged in 2002 (95% Bayesian credibility intervals 2001.0 to 2003.9) from a group of six Singaporean isolates. The Singaporean isolates were from between 2004 and 2007 and had divergence dates ranging from 1999 to 2002 (total range 95% credibility intervals 1998.1 to 2004.8). The SNP distance between the Hong Kong node and the closest Singaporean isolate was two SNPs therefore indicating that a transmission event from Singapore to Hong Kong between 1998 and 2004 was highly likely.

3.2.3.2 Guernsey

The 21 isolates from Guernsey demonstrated a mixed population of highly related and more sporadically distributed isolates.

Ten of the 21 isolates from Guernsey (43%) clustered to a single branch on the tree that diverged in 1998 (95% Bayesian credibility intervals 1996.5 to 2000.7) (Figure 10 and Figure 11). This cluster included seven of the twelve isolates sampled from the ITU at PEH, the only acute hospital trust on Guernsey, between 2006 to 2010 and had an average pairwise SNP distance of 7 SNPs (range 3 to 12). An additional three contemporaneous Guernsey isolates from 2014 to 2015 that were sampled from PEH inpatient wards (not ITU) also grouped to the branch but had had an average pairwise SNP distance of 87 SNPs (range 84 to 94) indicating they are less closely related to each other and to the ITU clustered isolates. The closest isolate to the branch containing the ten Guernsey isolates on the MCC tree (Figure 11) was a UK isolate sampled from London that was 23 SNPs from the Guernsey branch node.

The remaining 12 Guernsey sampled isolates did not cluster to the single branch described above. Five contemporaneously sampled isolates from PEH inpatient ward that were not linked in space and time grouped together to a single branch on the tree, with an average pairwise distance of 70 SNPs (range 6 to 102), next to two UK West Midlands (Coventry and Warwickshire) sampled isolates that were 24 SNPs from the nearest common ancestor. The remaining seven Guernsey isolates were sporadically distributed with UK genomes and showed no specific clustering with each other. This would indicate a population structure in Guernsey that comprises of a dominant local lineage alongside sporadic strains that are more related to UK mainland isolates.

3.2.3.3 Birmingham isolates and their relationship to UK/Global phylogeny

Out of the total 427 ciprofloxacin resistant genomes in the global phylogeny 94 grouped to a large clade on a single branch at the distal end of the tree. This clade contained isolates sampled between 2002 and 2015 from all three Birmingham hospital trusts and two isolates from Coventry (West Midlands, UK) (Figure 10). All isolates within the cluster appeared to belong to a single EMRSA-15 sub-lineage, due to their evolution from a single branch of the tree, which diverged in approximately 1995 (95% Bayesian credibility intervals 1994.1 to 1997.5). The average pairwise distance of all isolates in the clade was 72 SNPs, indicating a high degree of relatedness within the clade. This sub-lineage was subsequently named the BHM clade due to 98% of isolates within the clade being sampled from Birmingham and no genomes sampled from the rest of the UK or any other global sampling locations resolving to within the clade. Within the BHM clade there was evidence of further clustering with isolates being

distributed across two sub-lineages and evidence of localized evolution with epidemiologically linked isolates (section 3.2.4).

The 97 Birmingham genomes that did not group to the BHM clade were distributed throughout the rest of the tree amongst the UK and international genomes however small scale clustering events occurred. Eight clusters of between three and 17 Birmingham genomes that grouped to single branches were evident. The largest of cluster comprising of 17 genomes that were sampled between 2002 and 2015 grouped to a single branch that diverged in 1991 (95% Bayesian confidence intervals 1988.47 to 1994.97) positioned at the centre of the tree and was the largest grouping of Birmingham isolates outwith the BHM clade. Of these eight Birmingham clusters (outside of the BHM clade), three had West Midlands and five had UK sampled isolates basal to the single branches containing each of the clusters.

Thirty three Birmingham isolates (17%) were sporadically distributed throughout the tree and did not show major clustering or branching with other Birmingham isolates. Nineteen of these isolates had a West Midlands genome as the closest related isolate, nine had a UK isolate and five had a European sampled isolate as the closest related isolate.

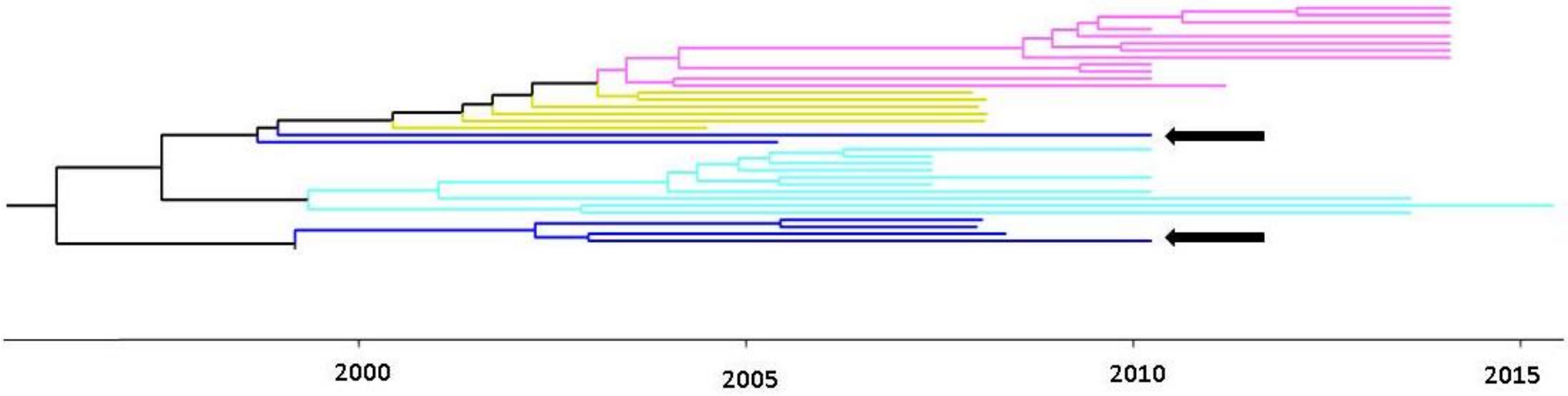

Figure 11: Branch of the MCC tree in Figure 6 containing the Hong Kong (pink), Singapore (yellow) and Guernsey cluster isolates (teal). Isolates from the UK are coloured dark blue and denoted with black arrows.

3.2.4 Birmingham Phylogenetic structure

An MCC phylogenetic tree consisting only of isolates from Birmingham (n = 188) was created with BEAST and coloured by sampling hospital (Figure 12). There was no evidence of hospital specific phylogenetic structuring with isolates from all three hospitals being distributed throughout the tree. Six small clustering events occurring between 2010 and 2015 (ranging between four and 11 isolates) that were positioned at the end of long single branches were evident. Three of these clusters were from the hospital with the greatest number of isolates (HEFT), two were from QEHB and a sixth cluster was formed of isolates from all three hospitals. Where possible epidemiological data linked isolates in these clusters in space and time. This included six isolates sampled from a period of increased incidence (PII) occurring on a stroke unit at HEFT over a period of 35 days (cluster 5 in Figure 12) and four screening isolates from a neurosurgery ward over 18 days from QEHB (cluster 4 in Figure 12). Notably the cluster of 17 isolates (cluster 2 in Figure 12) was not only the largest cluster but was the only post-2010 cluster to contain isolates from all three hospitals therefore indicating possible transmission between the Birmingham hospitals. The topology of these closely related clusters at the end of long single branches post 2010 may be indicative of more localized evolution of EMRSA-15 since the reduction in MRSA burden between 2001 to present in the UK.

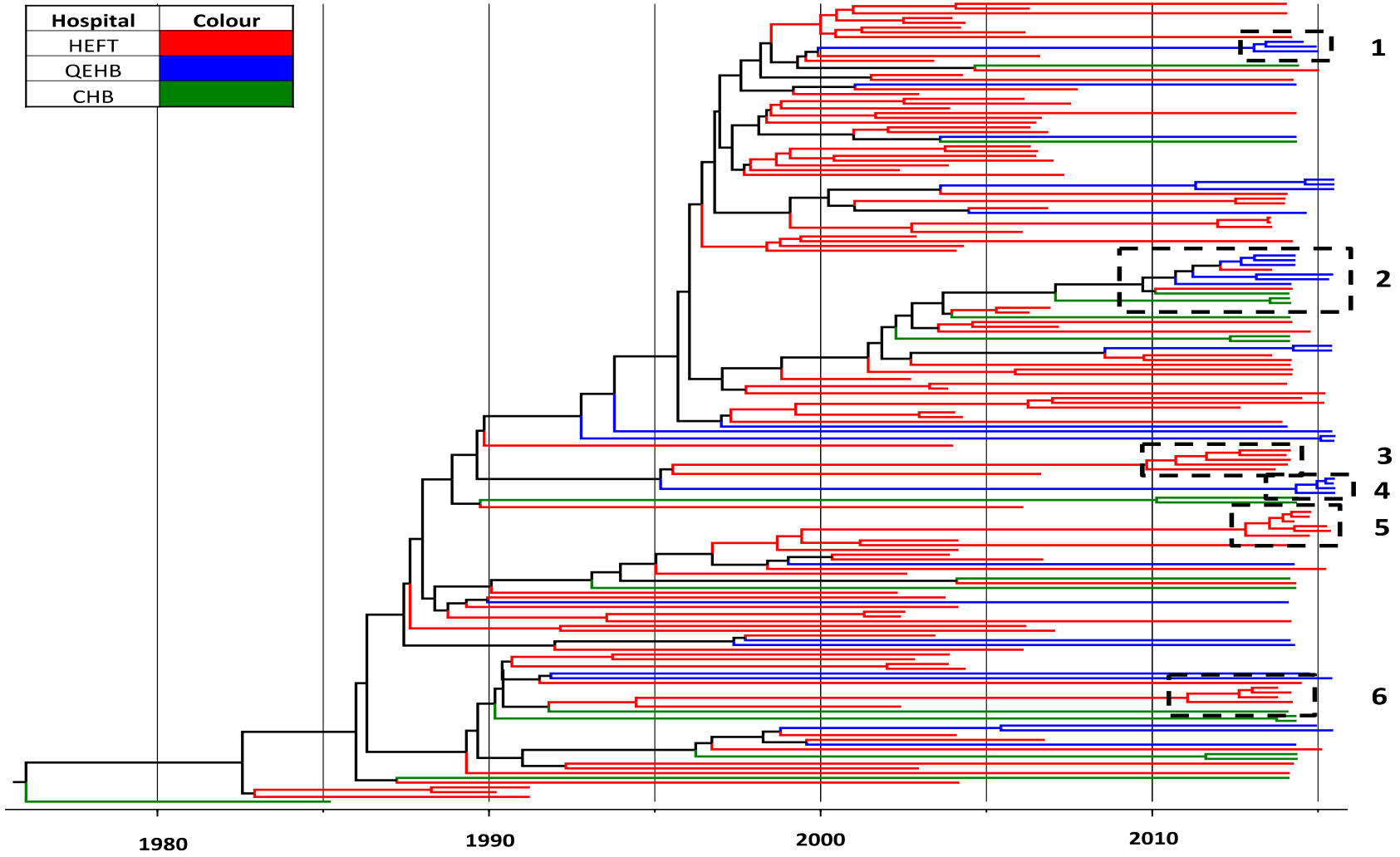

Figure 12: MCC tree of genomes from Birmingham hospitals only. Each branch tip is constrained by sampling date with the time scale displayed on the x axis. Branches are coloured by their sampling hospital and figure legend is displayed in the top left of the diagram. Clusters of isolates demonstrating localized evolution are displayed in dashed boxes.

The BHM clade contained 91 (52%) of the Birmingham sampled genomes and the constituency of the BHM clade changed with time. In 2002 only 17% of Birmingham genomes were found in this clade, whereas from 2003 onwards it constituted greater than 48% of sampled isolates for each year there were isolates available (no isolates were sourced from Birmingham between 2008 to 2012). Due to lack of isolates available from the QEHB or City Hospital pre dating 2012 the movement of the BHM strain between hospitals cannot be evaluated, however branches of QEHB or city isolates showed divergence dates in 2000 and 2002 respectively. These branches diverged from clusters of HEFT isolates suggesting spreading from HEFT to QEHB and CHB populations shortly after the strain emerged, indicating possible transmission from HEFT to QEHB and CHB during this time.

3.2.4.1 SNP analysis of BHM clade isolates

All of the BHM clade DNA sequences were scrutinized for SNPs and indels that were present in every BHM clade genome that may distinguish the clade from the rest of the EMRSA-15 population or may provide an evolutionary fitness advantage.

Twelve SNPs and two indels were identified that differentiated the BHM clade to all other EMRSA-15 genomes (Table 19a and 19b). Seven of these SNPs were unlikely to impact on the fitness of the BHM clade due to three of them being synonymous mutations, three were intergenic and another was in a pseudogene that does not code for a functional protein. Five of the 12 SNPs were non-synonymous however two were in hypothetical proteins and are therefore difficult to infer any change to ecological fitness. The three remaining SNPs were in the genes for malate:quinone oxidoreductase 1

(*mgo1*), putative amino acid permease (*pmh*) and the DNA supercoiling gene topoisomerase IV subunit B (*grlB*). The amino acid change in *mgo1* was from alanine to valine which are both nonpolar amino acids and therefore unlikely to change protein structure. The amino acid change in *pmh* was from arginine to glutamic acid, basic to acidic amino acids, and in *grlB* was from proline to serine, nonpolar to polar, suggesting both of these mutations could cause structural changes to the proteins they code for.

An additional five isolates from MRSA bacteraemias in the West Midlands (two from Coventry and three from Shrewsbury) located directly basal to the BHM clade and had nine of the 12 identified SNPs including the five non-synonymous mutations. These SNPs were however included as distinguishing features of the BHM clade due to the sampling location and basal topology of these West Midlands genomes on the phylogenetic tree (Figure 13).

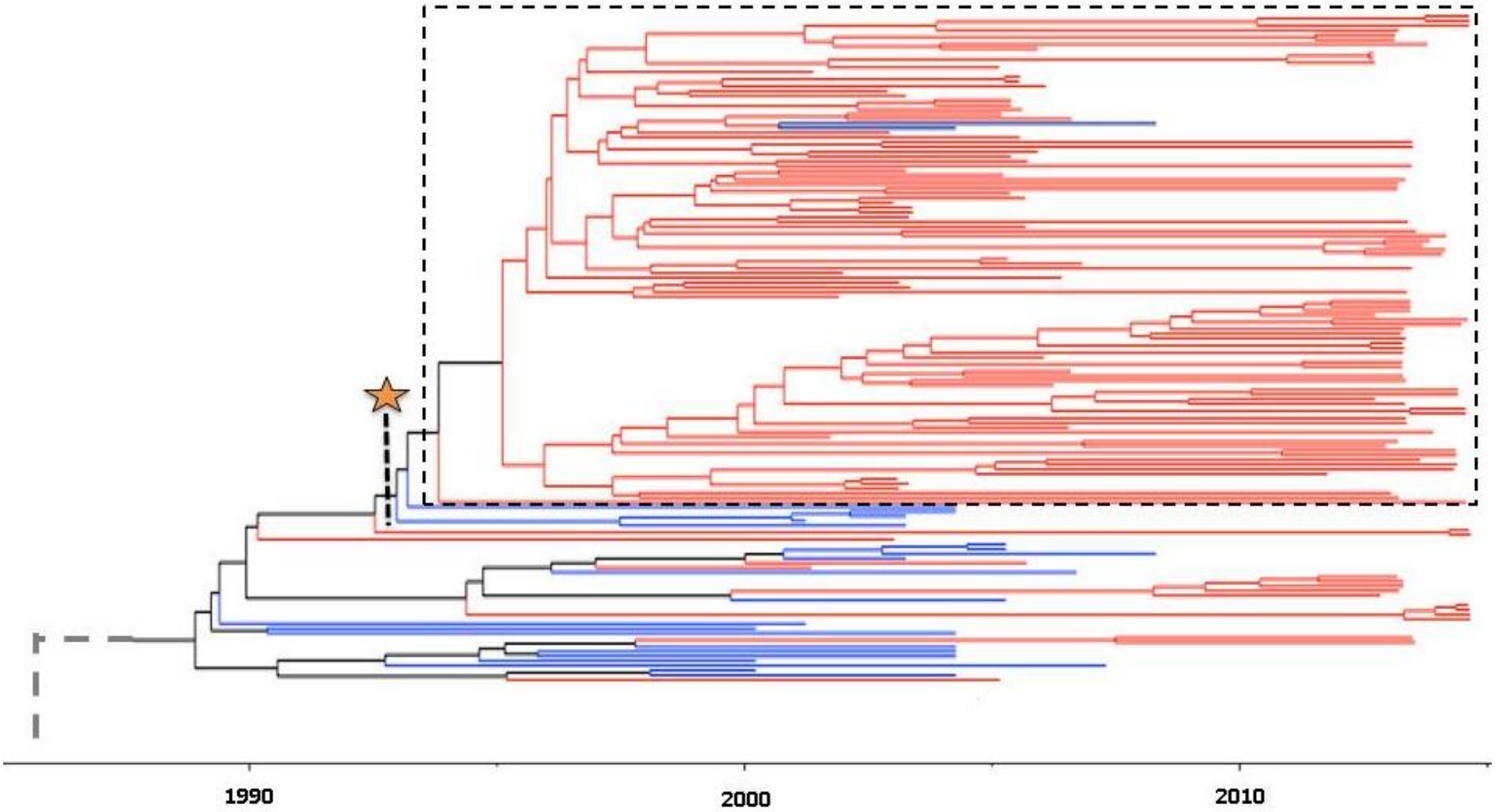

Figure 13: Enhanced region of the BHM clade and basal West Midlands isolates from the MCC tree in figure 5. The BHM clade is denoted in the dotted box. The evolution of the *grlB* mutation is depicted with a star and dotted line, with all isolates beyond this dotted cut off in the tree possessing the *grlB* mutation.

Position	SNP	Type	Amino acid change	Coding Sequence	GenBank ID
1184872	T to A	Intergenic	-	-	-
1184877	C to T	Intergenic	-	-	-
1334149	C to T	Non-synonymous	Pro 453 Ser	topoisomerase IV subunit B *(grlB)*	SAEMRSA15_12020
1433441	G to A	Pseudogene	-	ebh*	SAEMRSA15_12970
1679948	G to T	Intergenic	-	-	-
2352714	G to A	Synonymous	-	putative molybdenum transport system permease	SAEMRSA15_21760
2442618	G to A	Non-synonymous	Ala 132 Val	putative malate:quinone oxidoreductase 1 (*mgo1*)	SAEMRSA15_22660
2515404	G to A	Non-synonymous	Ala 50 Val	hypothetical protein	SAEMRSA15_23310
2538962	G to A	Non-synonymous	Arg 323 Glu	putative amino acid permease (*pmh*)	SAEMRSA15_23500
2732101	G to A	Synonymous	-	sensor kinase protein	SAEMRSA15_25280
2758770	C to T	Synonymous	-	hypothetical protein	SAEMRSA15_25490
2759407	G to A	Non-synonymous	Asp 300 Ile	hypothetical protein	SAEMRSA15_25490

Table 19a: SNPs that distinguish the BHM clade from the rest of the EMRSA-15 population. Genes located on the reverse DNA strand are in blue. * Protein is truncated and non-functional in EMRSA-15 due to non-sense mutation in gene (Holden *et al* 2013).

Position	Type	Location	Product	ID (Genbank)	Alternative	Reference
260331	Deletion	CDS	Putative PTS transport system, IIC component	SAEMRSA15_02060	TA	T
2820060	Insertion	CDS	ABC transporter ATP-binding protein	SAEMRSA15_26000	C	CA

Table 19b: Indels that distinguish the BHM clade from the rest of the EMRSA-15 population.

3.2.4.2 Pan genomic analysis

Nullarbor identified 328 genes present in all 91 BHM genomes, 69 of which were hypothetical. All of the remaining 259 genes were also identified in genomes out with the BHM clade, suggesting MGEs do not play a role in the success of the BHM lineage.

3.3 Comparison of EMRSA-15 diversity using WGS and VNTR

3.3.1 VNTR profiles

Forty two distinct VNTR profiles were identified from the 210 EMRSA-15 isolates. VNTR profile 1 2 0 3 17 – 2 was the most predominant (73 isolates, 34.8%) followed by 1 4 0 3 17 – 2 (38 isolates, 13.3%) and 1 3 0 3 17 – 2 (25 isolates, 11.9%). Twenty-nine VNTR profiles were only represented by single isolates.

3.3.2 Application of VNTR data to the Birmingham phylogeny

VNTR data for loci 13 and 21 were applied to the Birmingham MCC phylogeny (Figure 14). Locus 13 showed concordance with the MCC phylogeny in respect to the BHM clade. All isolates with 2 tandem repeats at loci 13 grouped to the BHM clade whereas isolates with 3,4 or 5 tandem repeats at locus L13 were distributed throughout the other branches of the tree. Therefore VNTR typing was able to identify the BHM clade with 100% concordance at locus 13. Locus 21 showed more variable mapping in relation to the Birmingham phylogeny and was concordant with WGS for closely related or known epidemiologically linked isolates. The first reported EMRSA-15 genomes were also

included in the phylogenetic tree, however no VNTR data could be generated for these genomes due to only the public genomes and not isolates being available for analysis.

In addition to the SNPs and indels in Table 19, all 91 isolates in the BHM clade were found to have a small unmapped region of 67 bases at the genomic location of VNTR locus 13 which was absent in all other genomes in the study, thereby further differentiating the BHM clade from the rest of the EMRSA-15 population.

3.4 SNP and VNTR comparative analysis

The 210 isolates that were WGS SNP typed and VNTR typed (Table 7) were analysed for relationships between SNPS, LVs and STRD. All versus all SNP distance was compared to the number of locus variants (LV) and STRD separately in Figure 15 and Figure 16 respectively. All versus all SNP distance showed a range from zero to 196 SNPs between paired isolates. No isolates varied at more than two loci (LV of 2) and STRD values ranging from zero to 14.

The number of SNPs differentiating each LV showed similar ranges and each LV value showing similar distribution of isolates with bigger circles (more isolates) between 50 and 100 SNP discrimination (Figure 15). SNP and STRD comparison showed a small decrease in range of SNPs that differentiated increasing STRD, with an STRD of one ranging from 0 to 196 SNPs and an STRD of 14 ranging from 14 to 150 SNPs (Figure 16).

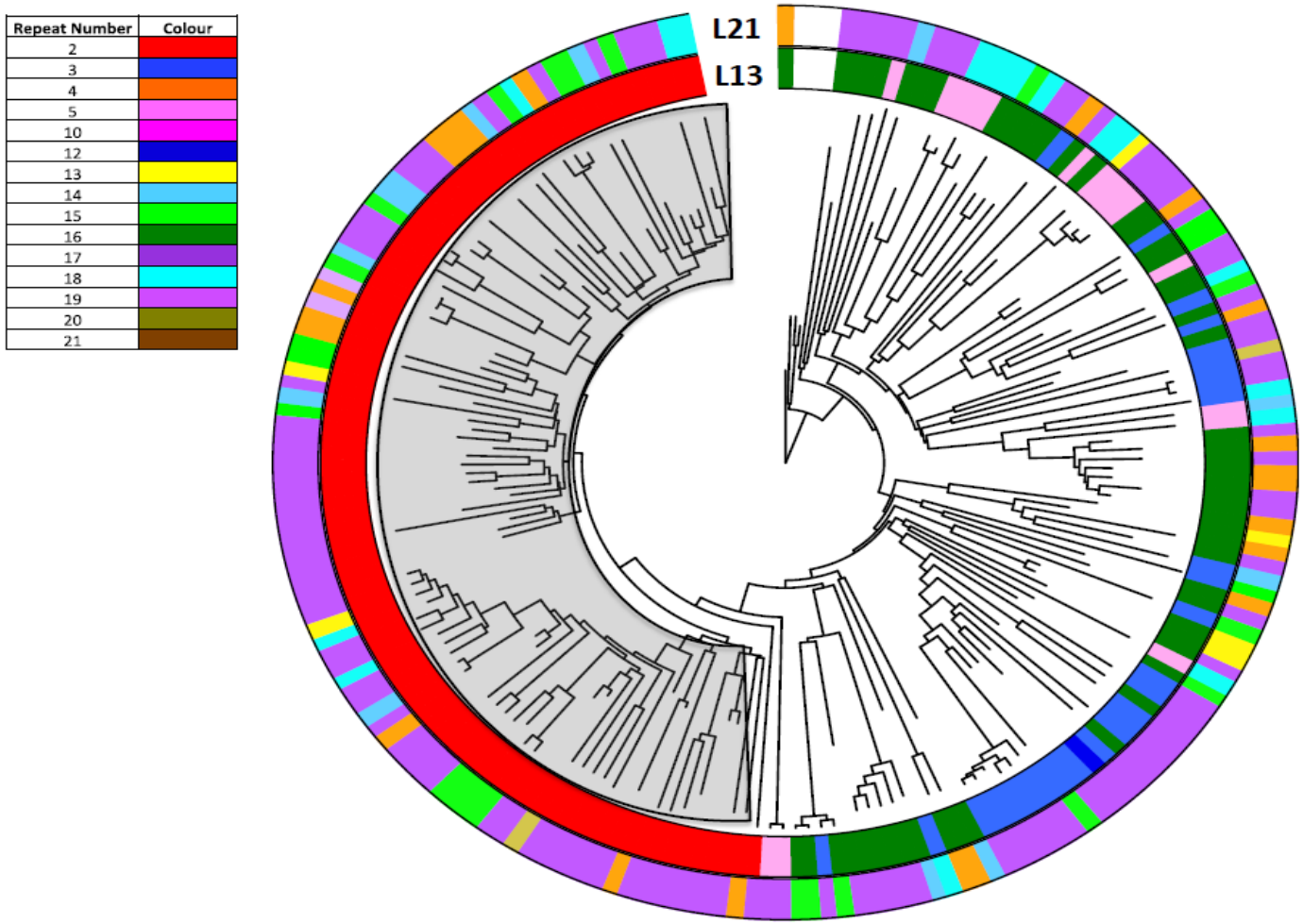

Figure 14: Circularized MCC tree of Birmingham EMRSA-15 isolates with the number of tandem repeats at VNTR locus L13 (inner ring) and locus L21 (outer ring) for each isolate concentrically displayed. Number of repeats are represented by different colours and described in the figure legend in the top right of diagram. The BHM clade is highlighted by a shaded box (left side of diagram). Branch lengths are not to scale.

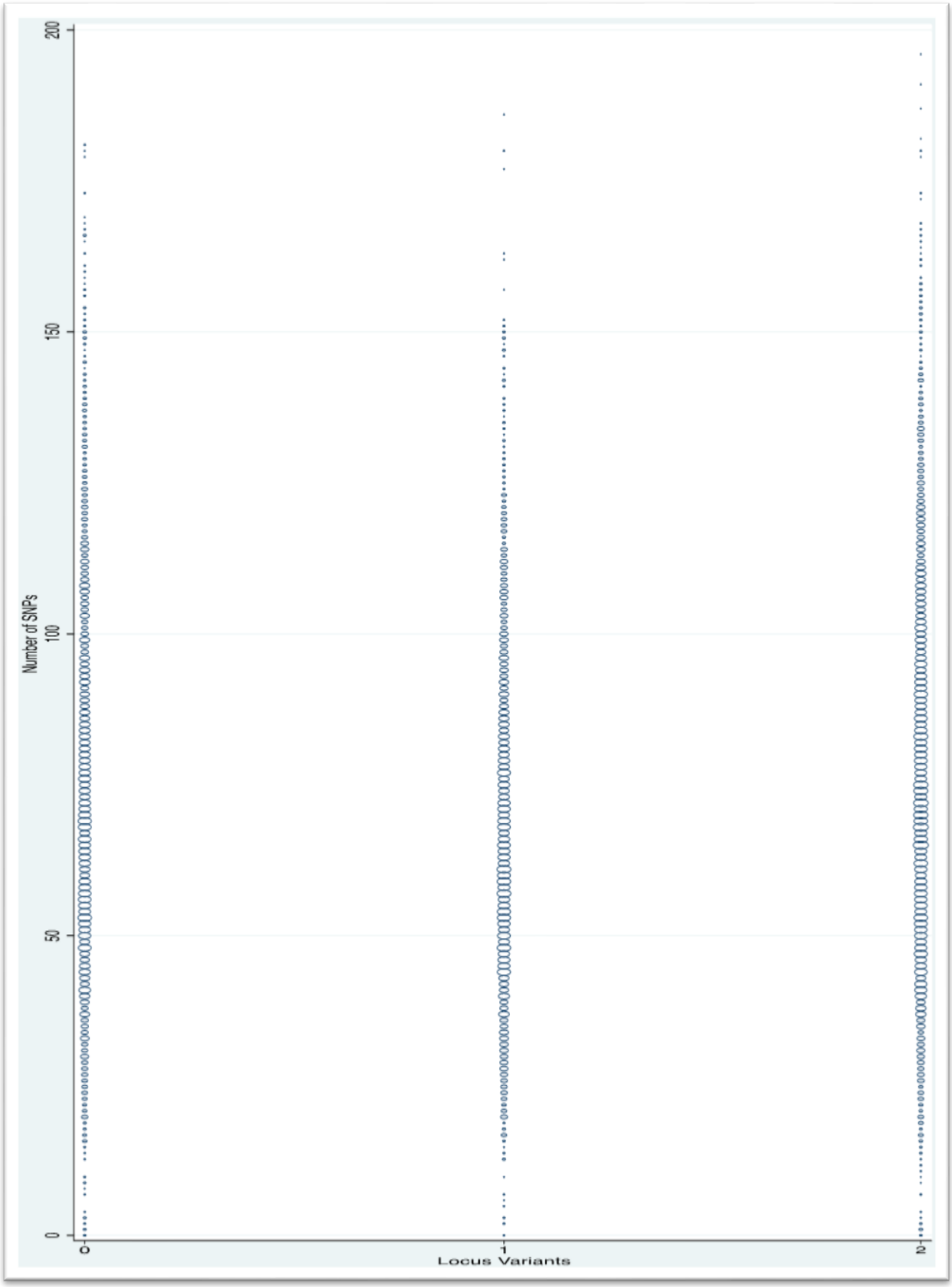

Figure 15: Bubble chart of SNPs between locus variants (LV) amongst the 210 WGS and VNTR typed EMRSA-15 isolates. Increasing bubble size represents increasing number of isolates.

Figure 16: Bubble chart of SNPs between STRDs amongst the 210 WGS and VNTR typed EMRSA-15 isolates. Increasing bubble size represents increasing number of isolates.

3.5 The relationship between phylogenetic structure and fluoroquinolone use in Birmingham

3.5.1 Antibiotic usage, occupied bed days and bacteraemia rate at HEFT

3.5.1.1 Occupied Bed days

The number of occupied bed days remained stable between 2001 and 2006 (SD 5431.18 OBDs) then rose by 175583 OBDs (46% increase) from 383077 OBD to 558660 OBD in 2007 due to the merger of Good Hope Hospital into the Heart of England Foundation trust (Figure 17). From 2007 to 2015 OBDs, continued to remain relatively stable (SD 21653.58) over time.

3.5.1.2 Antibiotic prescription at HEFT

Antibiotic prescription data was evaluated as defined daily doses (DDD) rather than DDD/OBDs due to the lack of OBD data predating 2001 and a stable OBD rate (excluding the inclusion of Good Hope Hospital). Fluoroquinolones, macrolides and 3GCs each had different usage patterns across the 25-year timeframe from 1990 to 2015. Between 1998 and 2008 fluoroquinolone drugs were the most frequently used of the three classes of antibiotic, being replaced by macrolides from 2008 onwards, while 3GC usage remained at a low level between 1991 and 2015. Individual drug use within each class varied and generated a different proportion of each drug class total (Figures 18 to 21).

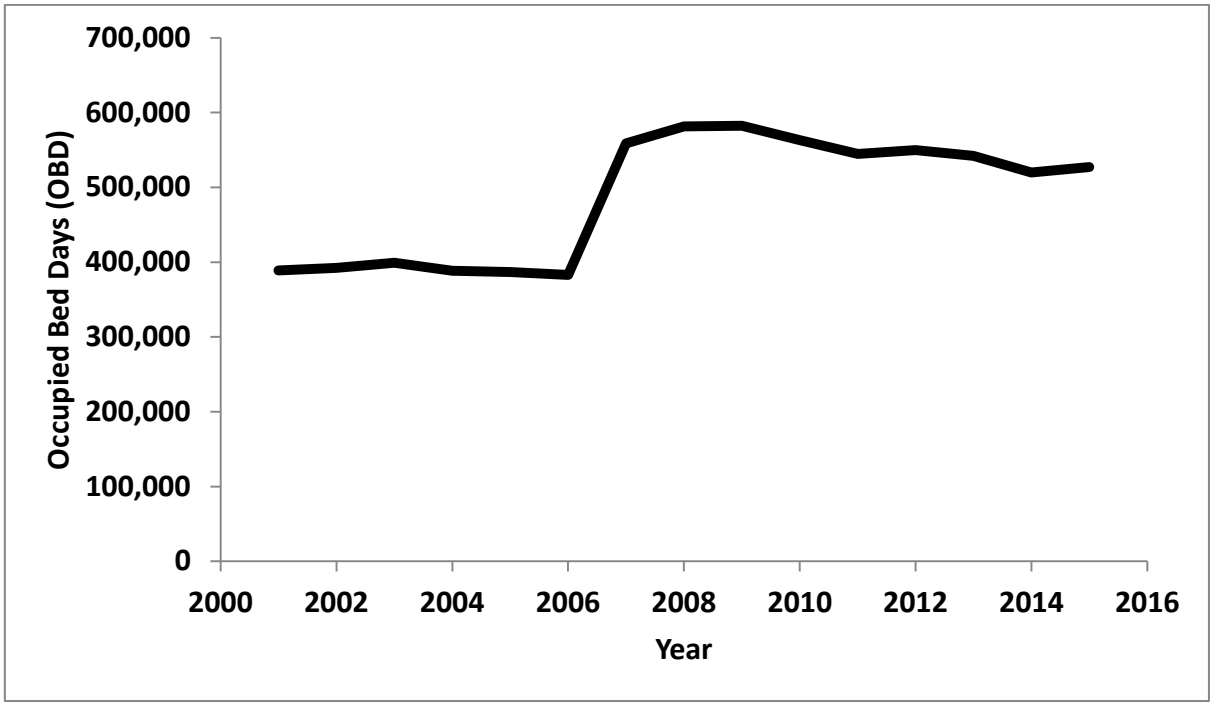

Figure 17: Occupied bed days (OBD) at Heart of England Foundation Trust between 2001 and 2015. * Rise in OBD in 2007 is attributed to the merger of Good Hope Hospital into the HEFT trust.

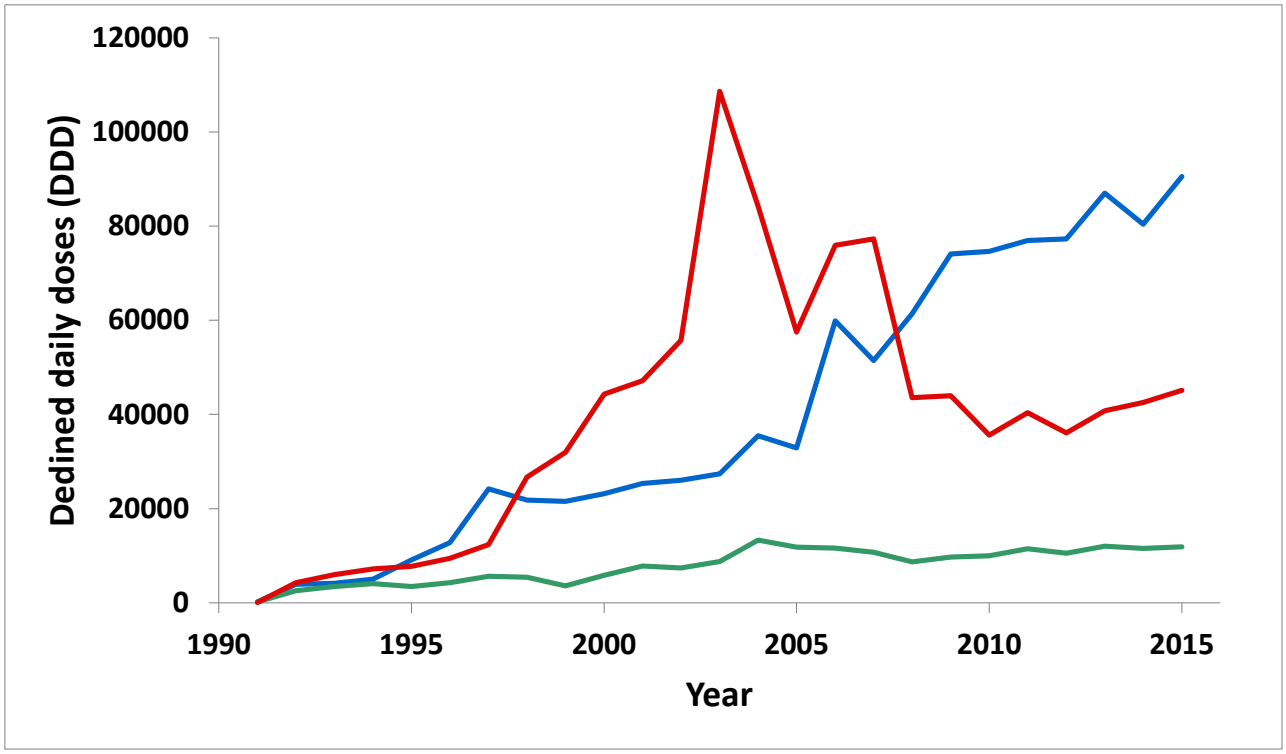

Figure 18: Defined daily doses of fluoroquinolones (red), macrolides (blue) and third generation cephalosporins (green) between 1990 and 2015 at HEFT.

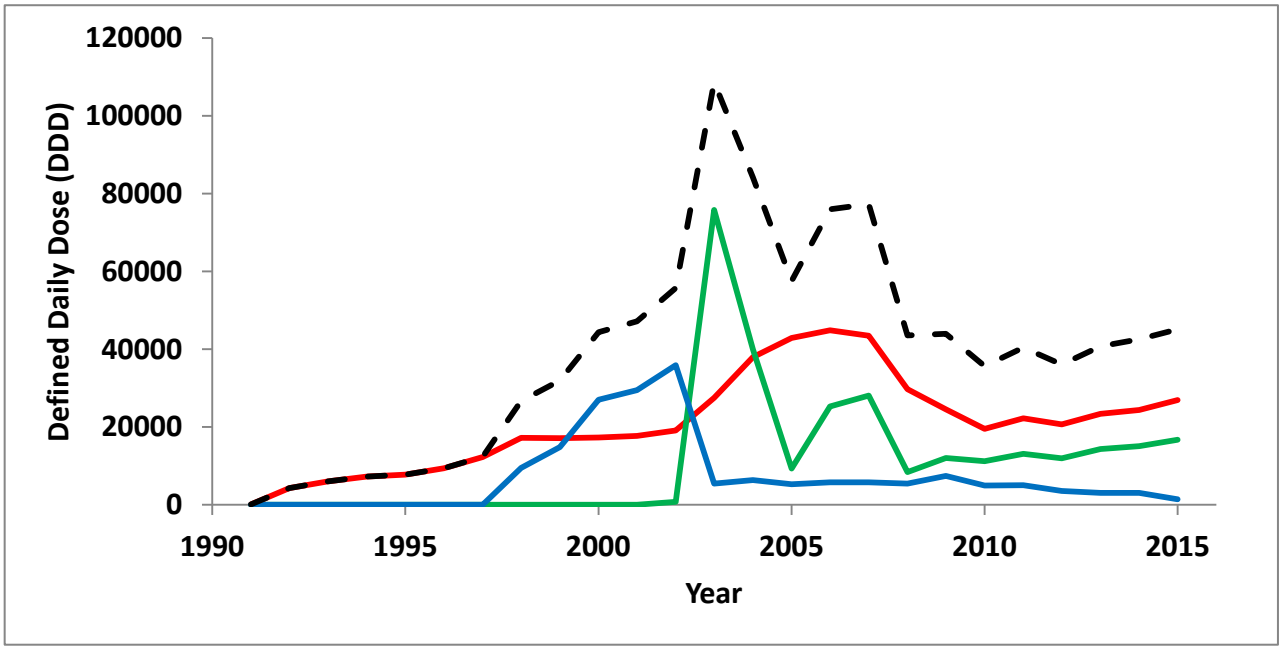

Figure 19: Use of ciprofloxacin (red), levofloxacin (green) and ofloxacin (blue) at HEFT between 1990 and 2015. Total fluoroquinolone use is represented by a dashed black line.

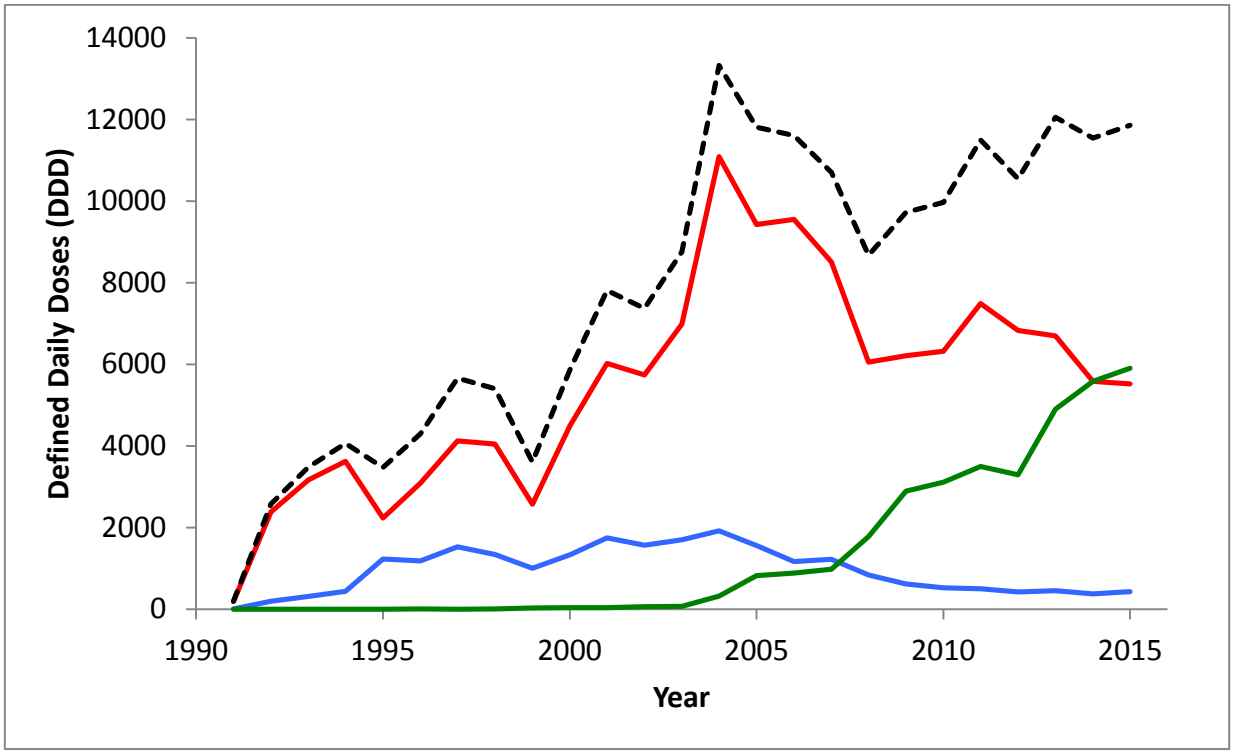

Figure 20: Use of ceftazidime (red), ceftriaxone(green) and cefotaxime (blue) at HEFT between 1990 and 2015. Total 3GC use is represented by a dashed black line.

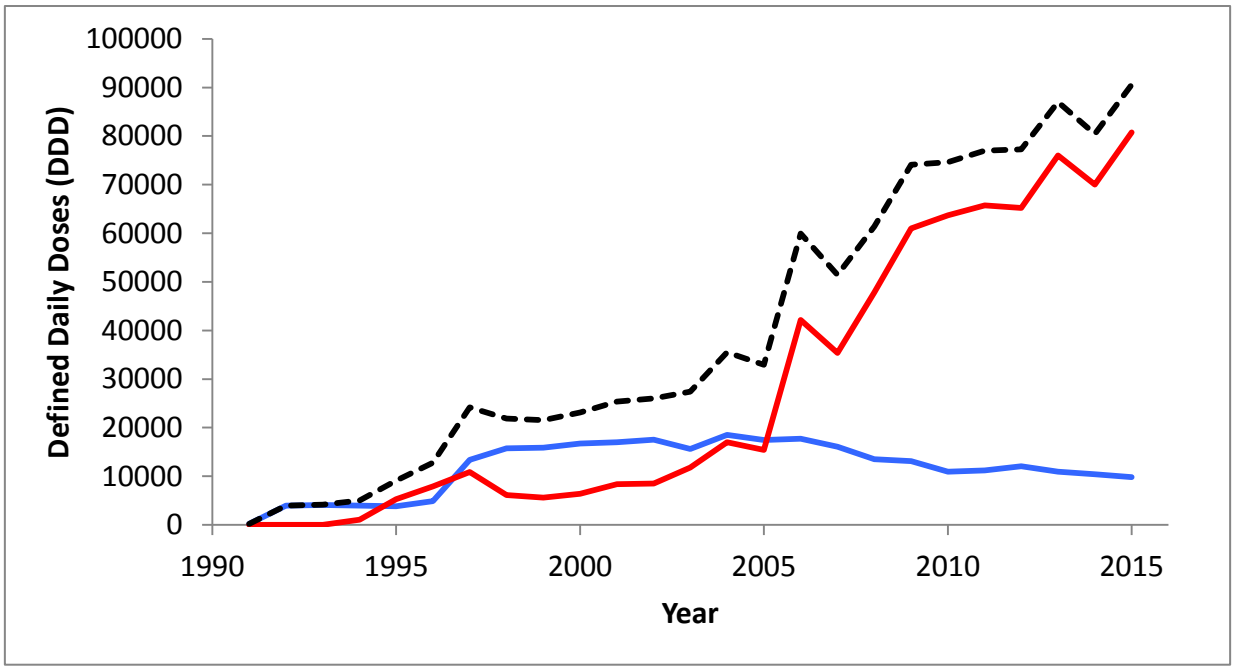

Figure 21: Use of clarithromycin (red) and erythromycin (blue) at HEFT between 1990 and 2015. Total macrolide use is represented by a dashed black line.

3.5.1.2.1 Fluoroquinolones

Total fluoroquinolone antibiotic usage showed a variable rate between 1990 and 2015 with rapid increase when the second generation drugs levofloxacin and ofloxacin became available (Figure 19).

In 1990 no quinolones were used in HEFT, with ciprofloxacin being introduced in 1991, ofloxacin in 1998 and levofloxacin in 2002. Ciprofloxacin usage increased gradually from 1991 to 1998 and then further increased to peak usage at 44859 DDD/year in 2006. Following 2006 there was a gradual decrease in ciprofloxacin usage until 2010. In contrast ofloxacin and levofloxacin usage increased sharply after their introduction to peak usage in 2002 and 2003 respectively. Changes in antibiotic prescribing policy were reflected in the fluoroquinolone antibiotic prescription data, with a switch from ofloxacin to levofloxacin in 2003 and reduced use of levofloxacin from 2004 to 2007. Between 2010 and 2015 total use of fluoroquinolones has gradually increased by an average of 3008 DDD/year due to increase in ciprofloxacin and levofloxacin (increasing annually by an average of 1485 and 1108 DD/year respectively). Ofloxacin use declined between 2010 and 2015 by an average of 694 DDD/year to an all time low usage of 1446 DDD/year in 2015.

3.5.1.2.2 Third generation cephalosporins

Use of third generation cephalosporins remained at relatively constant low usage between 1990 and 2015 (minimum and maximum of 196 DDD/year and 13324 DDD/year in 1991 and 2004 respectively) and had the least frequent usage of the three drug classes investigated (Figure 18 and Figure 20). Between 1990 and 2014

ceftazidime was the most frequently used cephalosporin. In 2015 ceftriaxone replaced ceftazidime as the most frequently used drug, having steadily increased in use since 2003. Over the 25 years cefotaxime remained at low usage, with a reducing trend from 2004 to 2015.

3.5.1.2.3 Macrolides

Macrolide usage increased annually from 225.5 DDD/year in 1991 to 90556 DDD/year in 2015 with an average increase of 3763 DD/year (Figure 21). This increase is mainly attributed to the annual increasing use of clarithromycin (average 3363 DDD/year) which rapidly increased in usage between 2005 and 2015, possible due to being part of the treatment algorithm for community acquired pneumonia.

3.5.1.3 MRSA bacteraemias

MRSA bacteraemia rate fluctuated over time and both mandatory and voluntary bacteraemias per year at HEFT are displayed in Figure 22. Voluntary data showed annual rates to primarily decreasing between 1998 and 1999, then increase again from 1999. Mandatory reporting was introduced in 2001 and showed a further increase in MRSA rate to a peak of 140 MRSA bacteraemias per year in 2003 followed by a reduction then spike again in 2006. From 2006 onwards there was a notable decline in bacteraemia rate until 2010 where the rate plateaued until 2015 when the lowest MRSA rate was recorded at 13 bacteraemias per year.

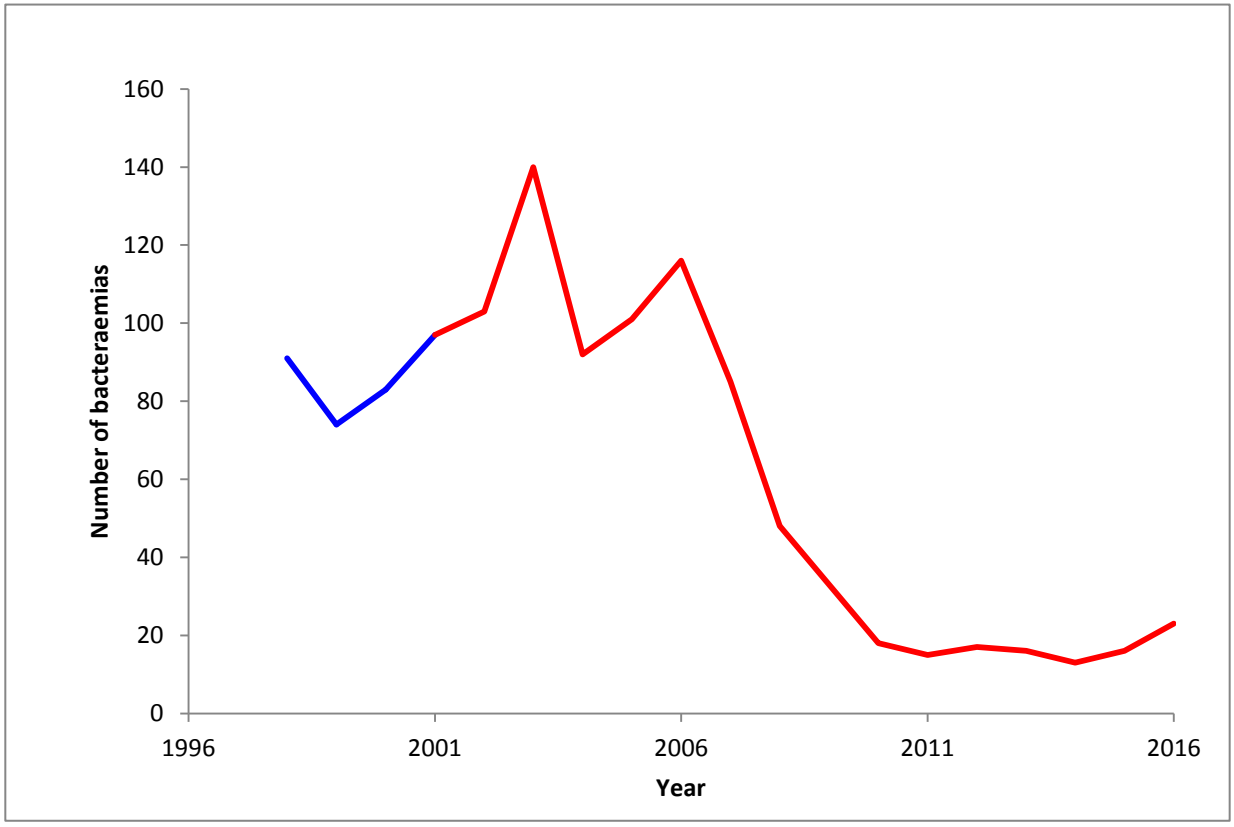

Figure 22:Annual MRSA bacteraemia rate at HEFT. Mandatory reporting was introduced in 2001 and is displayed in red. Data from before 2001 was voluntarily reported and displayed in blue.

3.5.2 Impact of fluoroquinolone use on the EMRSA-15 phylogeny in Birmingham over time

In order to further understand the evolution of EMRSA-15 in Birmingham the impact of fluoroquinolone use on the phylogenetic structure of the Birmingham MCC tree was investigated by pairing the time linked Birmingham phylogeny with HEFT annual fluoroquinolone data (Figures 22 and 23). Only fluoroquinolone prescription and not 3GC or macrolide prescription was chosen for analysis due to a concordant pattern and a positive Pearson's correlation of 0.65 and being obtained between annual MRSA bacteraemia and total fluoroquinolone usage at HEFT (Figure 23). Negative correlations of -0.32 and -0.84 were observed between bacteraemias and total use of third generation cephalosporins and macrolides respectively.

The EMRSA-15 population in Birmingham demonstrated a linked relationship of increasing diversity with the increasing use of fluoroquinolones over the 25 year time period but has also established that the early evolution of EMRSA-15 occurred before fluoroquinolones were introduced at HEFT.

Figures 23 and 24 demonstrate that the early evolution of EMRSA-15 occurred before the introduction of ciprofloxacin in 1991. Our global MCC tree has demonstrated EMRSA-15 to have emerged in 1984 (Figure 10) which was seven years before ciprofloxacin was introduced at HEFT (Figure 23). Further evidence of pre-ciprofloxacin use evolution can be seen with major branches of the tree developing between 1985 and 1990 (Figure 14), which continued to evolve post 1991, and account for a large proportion of the diversity within EMRSA-15 population in Birmingham.

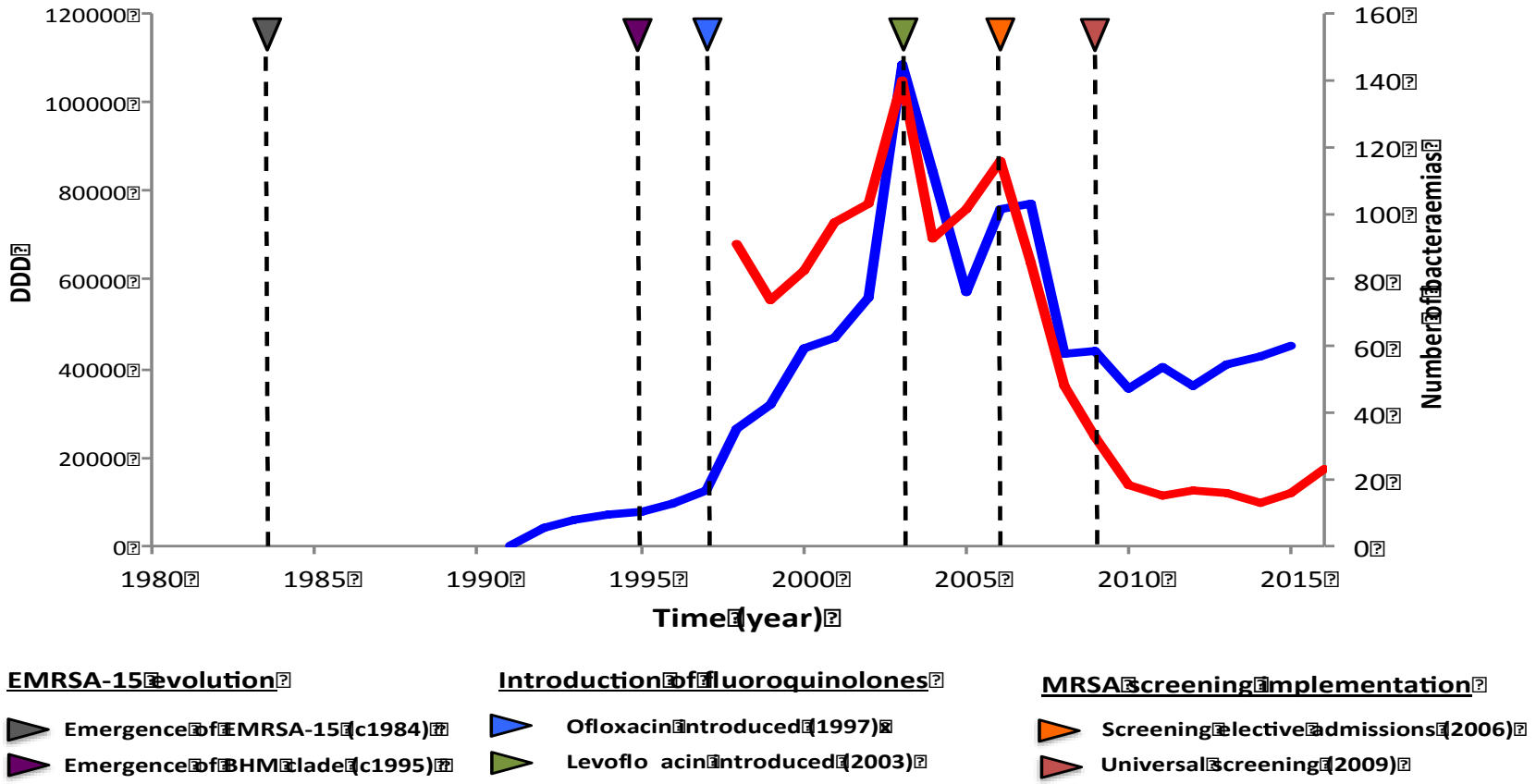

Figure 23: Graph of total fluoroquinolone use (left y axis [blue]) and annual bacteraemias (right y axis [red]) at HEFT. Key dates for EMRSA-15 evolution, new fluoroquinolone drug introduction and IPC screening policy have been annotated with coloured triangles and are detailed below the graph.

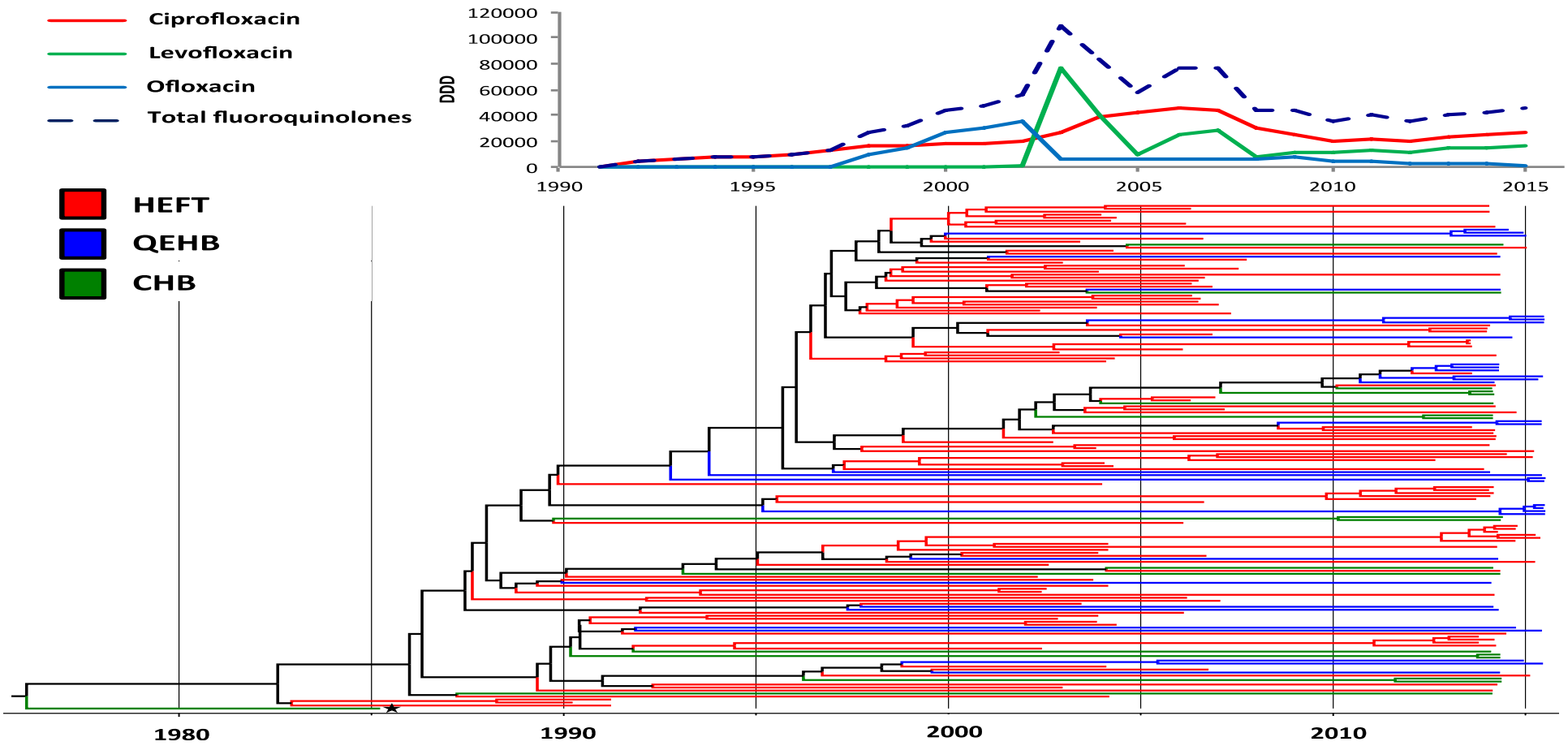

Figure 24: Maximum clade credibility (MCC) tree of EMRSA-15 in Birmingham between 1985 and 2015 coloured by sampling hospital. Branch tips are constrained by sampling time and time scale is displayed along the bottom (x axis). Fluoroquinolone antibiotic use per year between 1990 and 2015 is displayed above the tree. Colour legend for antibiotic usage is located top left of figure and the colour legend for the MCC tree is located below with coloured boxes. The historical Birmingham isolate is denoted with a star.

Between 1991 and 1996 ciprofloxacin use increased from 0 to 9000 DDDs/year, increasing by approximately 2000 DDD/year. During this time the diversity of the EMRSA-15 population increased as demonstrated in 21 branching events occurring between 1991 and 1996 (including the emergence of the BHM lineage containing the *grlB* mutation) compared to 11 branching events before 1991 when ciprofloxacin was not being prescribed.

From 1996, when ofloxacin was introduced, to 2003, when fluoroquinolone usage peaked after the introduction of levofloxacin, there was a 4.5 fold increase in total ciprofloxacin usage and three fold increase in EMRSA-15 diversity with 59 branching events on the phylogenetic tree, 46 of which occurred within the BHM lineage, over this seven year period.

Between 2003 and 2010 fluoroquinolone use dropped to an all time low rate and branching of the tree was reduced to 21 events (18 within the BHM lineage). During this time local and national MRSA hospital infection rates were falling (from the introduction of IPC) which may also be accounting for the reduction in branching.

Post 2010 the fluoroquinolone rate plateaued and then fluctuated but showed no drastic change however 35 branching events occurred between 2010 and 2015 suggesting localized evolutionary events. As described in section 3.2.4 this may be reflecting the reduced MRSA burden (Figure 22) and increased number of outbreak or PII isolates that were included in our isolate collection during this time.

The effect of IPC policies such as screening of hospital admissions (Figure 23), hand hygiene compliance may also have effected the phylogeny of EMRSA-15 in Birmingham, however analysis of these impacts are out with the scope of this study.

3.6 Desiccation tolerance of EMRSA-15

3.6.1 Validation of desiccation chamber

The desiccation assay was validated for coverslip sterility, temperature variability and humidity variability as these factors could affect the reliability of results.

3.6.1.1 Sterilization of glass coverslips

Quintuple glass coverslips that had been sterilized by either washing with absolute ethanol or UV irradiation at 450nm were incubated in BHI broth to test for sterilization efficiency. No growth in BHI broth was observed for any of the UV irradiated or ethanol washed coverslips or when 100µl inoculations of broth were plated onto CBA at 24 and 48 hours incubation. Quintuple coverslips that had no sterilization applied to them each had cloudy BHI broth culture and a lawn of bacterial growth on CBA at 24 hours. A negative control of BHI broth with no coverslips showed no cloudiness or evidence of bacterial growth at 24 or 48 hours. From these data UV irradiation was chosen to be the method for sterilization of coverslips for the desiccation assay.

3.6.1.2 Temperature and humidity control

The temperature and RH of a single ward at HEFT was tested on three separate days at approximately the same time each day (circa 1pm). Temperature remained stable at 23.0^{o}C and RH was seen to range from 36% to 46%.

Temperature in the desiccation chamber varied little over the validation time ranging 20^{o}C to 23.4 ^{o}C (SD 0.77) with an average temperature of 21.2 ^{o}C (Table 18). The highest

temperature value of 23.4^{o}C was recorded when direct sunlight was shining on the desiccation chamber. The desiccation chamber was then moved to avoid direct sunlight. This demonstrated the temperature in the lab was stable and additional temperature control of the desiccation assay would not be required.

RH was recovered over the same validation period of eight hours each day for three days but showed greater variation than temperature and ranged from 26.2% to 43.4% (SD 5.54) (Table 19) with an average of 36.8%.

Maintenance of a stable RH was tested with the addition of a saturated calcium chloride solution in the desiccation chamber and recording of RH every hour over eight hours in a single day and overnight (+ 16 hours). The RH in the desiccation chamber with the addition of the calcium chloride solution showed a more stable RH range of 32.7% to 33.2% (SD 0.14) compared to when no salt solution was added (Table 19).

Day	Time								
	9am	10am	11am	12pm	1pm	2pm	3pm	4pm	5pm
1	20.0	20.1	20.4	20.4	20.4	20.7	20.7	20.7	20.5
2	20.6	21.2	21.6	21.6	21.6	21.8	23.4	21.1	21.1
3	21.5	21.6	21.5	21.4	21.7	21.8	21.4	21.5	21.4

Table 20: Temperature record of desiccation box over three days.

Condition	Time									
	9am	10am	11am	12pm	1pm	2pm	3pm	4pm	5pm	9am (+16h)
No salt solution	39.6	39.3	35.6	38.1	37.4	34.8	31.9	30.4	32	32.7
Saturated $CaCl_2$ solution	39*	34	33.9	33.0	32.9	32.7	33	33.2	33	30.8

Table 21: Validation of desiccation chamber RH and CaCl2 salt solutions. * $CaCl_2$ solution was added at 9am.

3.6.2 Desiccation tolerance of EMRSA-15 strains suspended in water

Bacterial counts for all isolates suspended in water decreased over time, and after 28 days no viable bacteria could be detected (Figure 25 and 26).

No individual isolate demonstrated an overall ability to tolerate desiccation better compared to other isolates when suspended in water, with isolates that initially had a greater percentage survival in the first 240 hours showing poorer growth post 240 hours and vice versa. Rapid reduction in enumerable bacteria was observed in all 11 isolates at 12 hours post inoculation, however, this varied between isolates and ranged from 74.8% (HK_N5) to 15.7% (A901)(Figure 26). By 72 hours all isolates had dropped to below 3% survival and by 240 hours all isolates had dropped to below 1% survival. No growth was observed at 840 hours for any of the isolates; therefore the maximum detectable survival time for isolates suspended in water was 28 days.

Area under the curve (AUC) values for each isolate were calculated from percentage survival graphs and detailed in Table 22. Isolate HK_N5 had the highest AUC and the flattest linear regression slope for all isolates in the water assay (Figure 27 and Table 22). Isolate CT300 had the lowest AUC value the steepest linear regression slope when suspended in water. CT 300 also consistently had the lowest number of surviving bacteria at each time point of the assay, with the exception of 744 hours (where CT124 had the lowest count).

The ciprofloxacin sensitive EMRSA-15 progenitor isolate CT124 showed similar desiccation tolerance to the EMRSA-15 isolates when suspended in water. CT124 had the second highest AUC values with percentage survival values in the upper quartile range of values until 336 hours when viable counts began to rapidly drop. At 672 hours

(the last time point bacteria were detected at) CT124 had the lowest percentage surviving bacteria.

BHM clade isolates displayed higher AUC values compared to non-BHM isolates when suspended in water, with the three BHM clade isolate (CT147, ECMM195, A120) AUC values ranging from 1627.90 to 2234.72 (average 1900.67) versus the three non-BHM isolate (CT300, EMM513, A901) AUC values ranging from 1028 to 1689 (average 1304.33).

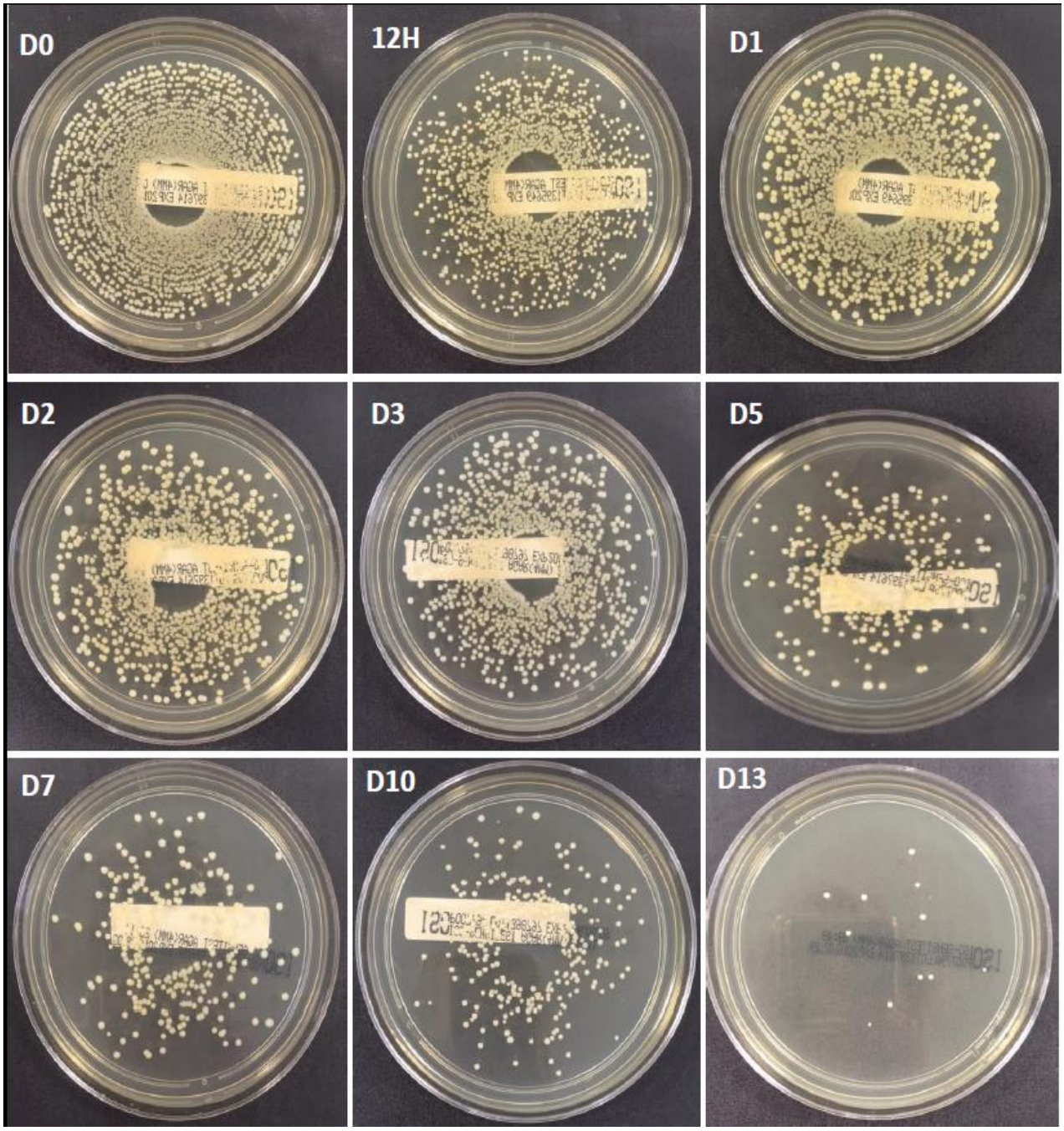

Figure 25: Spiraled plates of water suspended EMRSA-15 control isolate from day 1 to day 13. Plates after day 13 looked similar to the day 13 plate and are not shown.

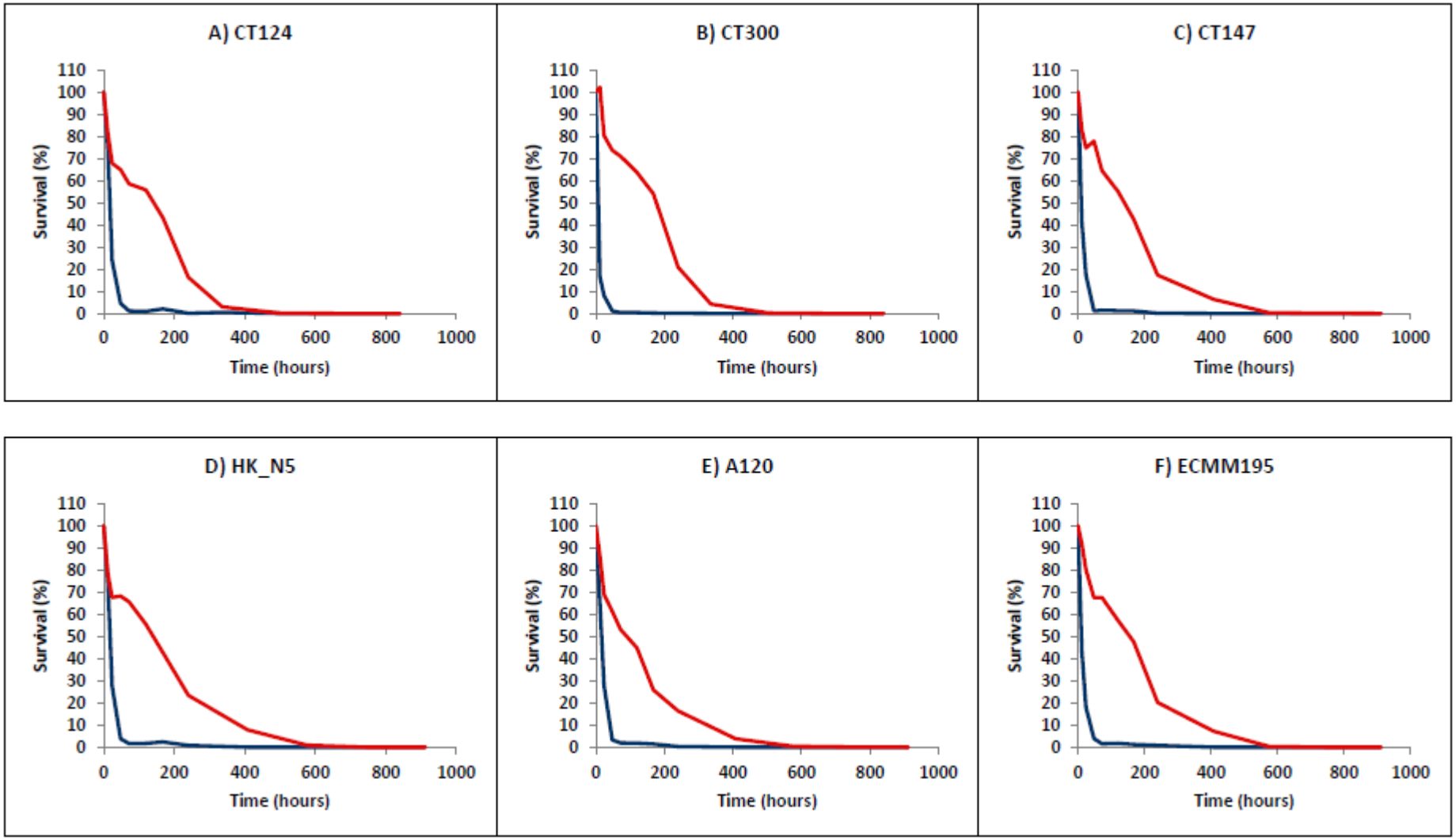

Figure 26a: Percentage survival of each of the 11 isolates in the assay. Water suspension assays are in blue and BSA suspension assays are in red.

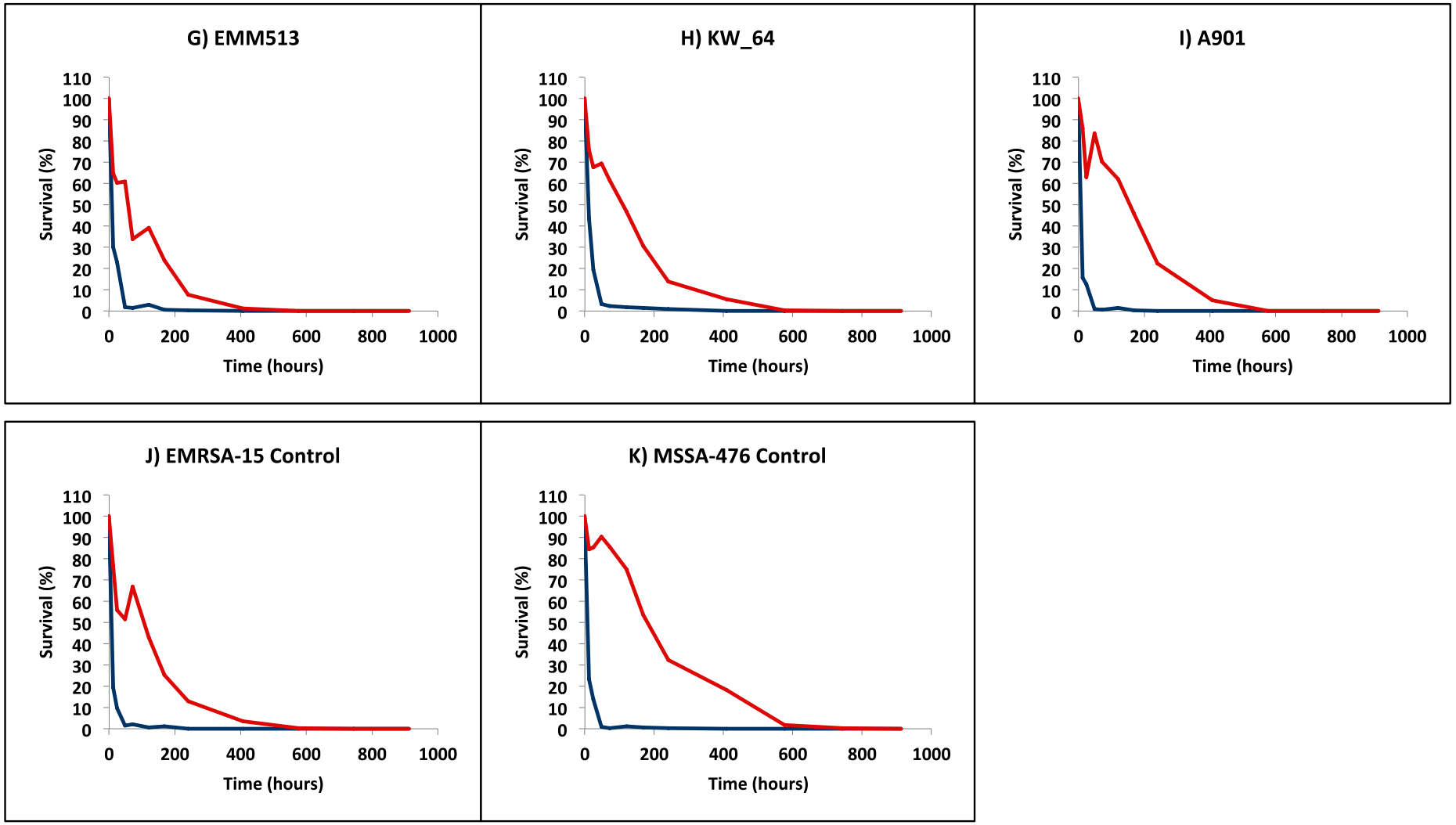

Figure 26b: Percentage survival of each of the 11 isolates in the assay continued. Water suspension assays are in blue and BSA suspension assays are in red.

Isolate	Area under the curve	
	Water	BSA
CT124	2262.26	14197.69
CT300	1028.52	17126.56
CT147	1627.80	15536.79
HK_N5	2462.54	16254.12
E15-C	1240.00	11929.56
MSSA-476	1307.37	22530.17
A120	2234.72	12646.99
ECMM195	1841.27	16485.82
EMM513	1689.73	9615.38
KW-64	1916.37	13373.32
A901	1196.97	16637.06

Table 22: Area under the curve (AUC) values from Figure 26 for each isolate suspended in water or BSA.

3.6.3 Effect of protein suspension menstrua on desiccation survival

All bacterial isolates suspended in BSA had an statistically significant increased in tolerance to desiccation compared to being suspended in water at every time point of the assay with the exception of isolate EMM513 at 840 hours (Table 21). Heavy growth on agar plates was visible until 168 hours (Figure 27) and the percentage of surviving bacteria remained above 50% until 72 hours post inoculation in all isolates (Figure 26). The percentage survival of all strains was considerably higher at 12 hours desiccation compared to water suspended isolates with a mean of 82.8% survival (range 102% to 64%); a two-fold increase in survival at this time point compared to water suspended isolates (mean 37.3%). All isolates also displayed between 5.7 and 17.2 fold increases in AUC values when suspended in BSA (Table 20) and all isolates suspended in BSA had linear regression lines with reduced slope gradient (Table 24) compared to water. Ten of the 11 isolates showed an increase in maximum survival time to greater than 35 days when suspended in BSA; a 25% increase in survival duration. Isolate CT300 showed a maximum survival time of 28 days, which was the same duration as when suspended in water.

Unlike when suspended in water, the control isolate MSSA-476 showed the highest desiccation tolerance when suspended in BSA due to having the highest percentage survival of bacteria at every time point in the assay and the highest AUC value. MSSA-476 had 3.5 fold more bacteria surviving than the EMRSA-15 isolate with the highest percentage survival (KW_64) at 840 hours (the last time point for bacteria to be detected) as well as the flattest slope gradient of all isolates in the assay.

Time	Isolate										
	CT124	CT300	CT147	HK-N5	EMRSA-15	MSSA-476	A120	ECMM195	EMM513	KW64	A901
0	3.05E-09	5.99E-08	2.33E-08	2.00E-05	6.34E-10	6.89E-07	5.66E-04	1.11E-07	6.64E-11	9.04E-14	2.66E-08
12	4.43E-07	8.88E-16	7.92E-10	1.52E-09	2.60E-15	2.61E-11	1.99E-04	4.80E-10	6.81E-12	1.78E-11	9.38E-18
24	1.22E-12	5.65E-09	6.35E-12	1.13E-11	2.44E-11	2.43E-07	1.60E-11	7.59E-15	9.91E-09	2.60E-10	1.82E-09
48	4.32E-13	2.69E-13	8.95E-13	7.88E-18	7.07E-11	2.10E-13	1.07E-12	3.76E-13	2.67E-10	9.89E-14	2.65E-15
72	1.41E-13	2.69E-13	1.44E-16	1.38E-14	1.67E-14	3.44E-21	4.72E-09	3.71E-15	1.53E-14	1.25E-16	7.73E-16
120	1.80E-15	6.40E-12	2.12E-14	5.64E-11	6.91E-10	3.89E-13	4.86E-06	3.29E-16	8.77E-11	1.25E-13	7.24E-15
168	8.05E-13	1.67E-13	4.23E-06	1.51E-11	1.44E-10	7.00E-11	5.52E-05	1.26E-13	1.40E-13	1.90E-10	2.04E-16
240	1.719E-12	6.68E-13	4.23E-07	5.07E-09	4.36E-08	9.95E-13	1.27E-07	2.57E-10	1.40E-10	9.00E-07	4.04E-12
336	1.36E-11	8.59E-10	1.84E-11	6.58E-14	5.79E-10	1.84E-14	5.52E-05	1.72E-15	2.91E-08	1.02E-09	1.13E-13
504	1.46E-06	1.10E-04	1.13E-07	3.22E-12	2.41E-04	2.86E-07	9.53E-11	9.20E-08	2.41E-05	3.29E-09	7.70E-09
672	1.12E-02	8.86E-03	7.00E-07	1.24E-06	8.38E-06	1.14E-05	6.70E-10	3.63E-11	1.96E-04	7.92E-11	3.92E-09
840	2.23E-02	-	1.83E-07	4.85E-03	1.03E-03	4.75E-14	5.47E-08	8.10E-06	7.60E-02	1.35E-07	1.55E-03

Table 23: P-values of statistical significant difference between water and BSA suspensions for each isolate and every time point of the desiccation assay. Values in white are below 0.001 statistical significance, blue are below 0.05 statistical significance and values in red show no statistical significance.

Isolate A901 displayed the highest desiccation tolerance amongst EMRSA-15 isolates and isolate EMM513 displayed the lowest desiccation tolerance, instead of CT300 in the water assay, when suspended in BSA. Isolate CT300 showed comparable tolerance to other EMRSA-15 isolates but was the only isolate to have no growth at day 840 hours.

CT124 also showed similar desiccation tolerance to ciprofloxacin resistant isolates when suspended in BSA. CT124 again showed rapid reduction in viable counts after 336 hours and had the lowest percentage surviving bacteria at the end time point of 840 hours.

Both BHM and non BHM isolates had a greater range of AUC values when suspended in BSA (Figure 26 and Table 20) however unlike when suspended in water these ranges over overlapped each other. Non-BHM isolates demonstrated both the lowest (EMM513) and highest (A901) EMRSA-15 isolate AUC values when suspended in BSA (range 9614 to 16637). BHM clade isolates had a reduced range compared to non BHM isolates (range 12646 to 16584) however averages for both BHM and non BHM clades were similar (14889 and 13792 respectively).

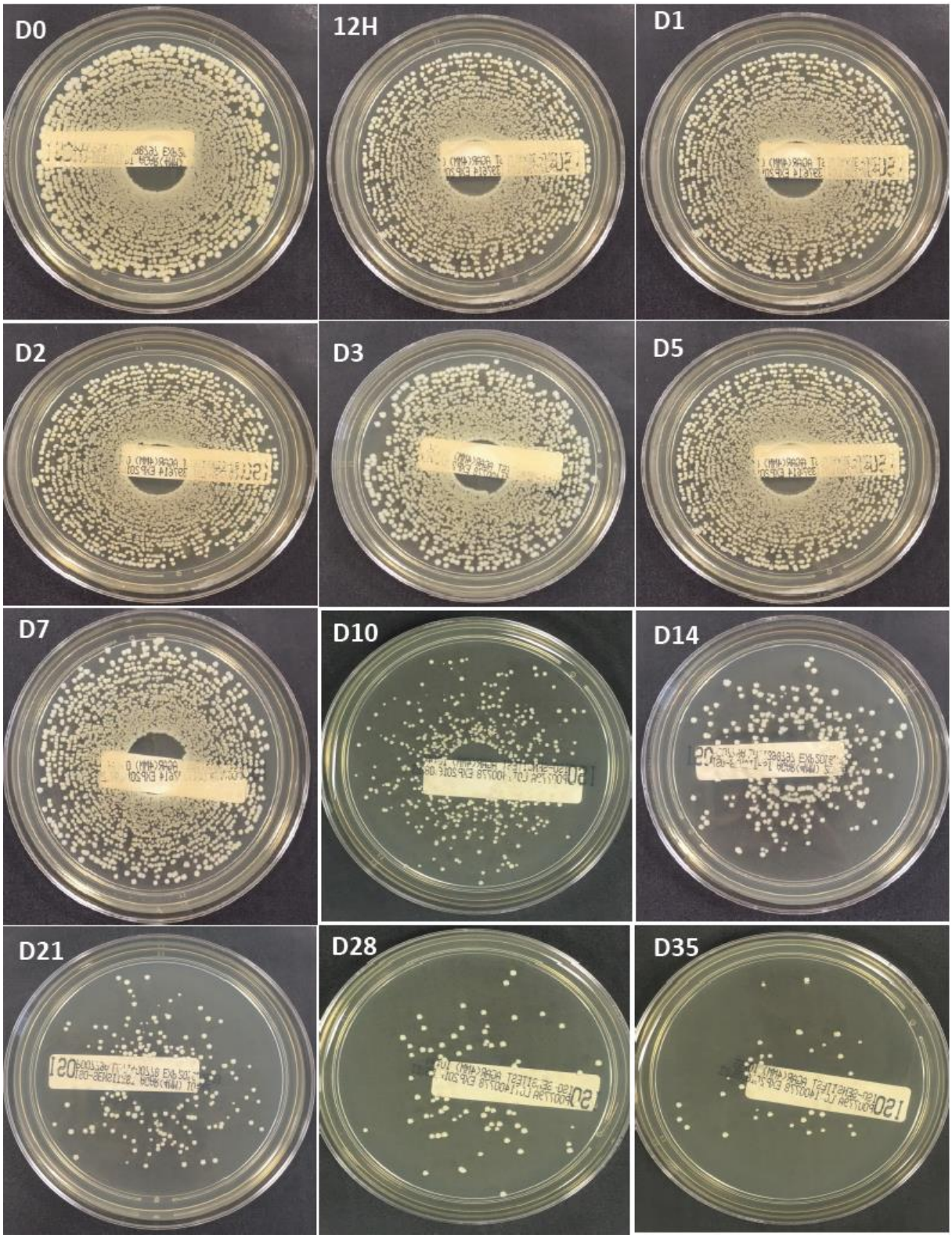

Figure 27: Spiraled plates of BSA suspended EMRSA-15 control isolate from day 1 to day 35.

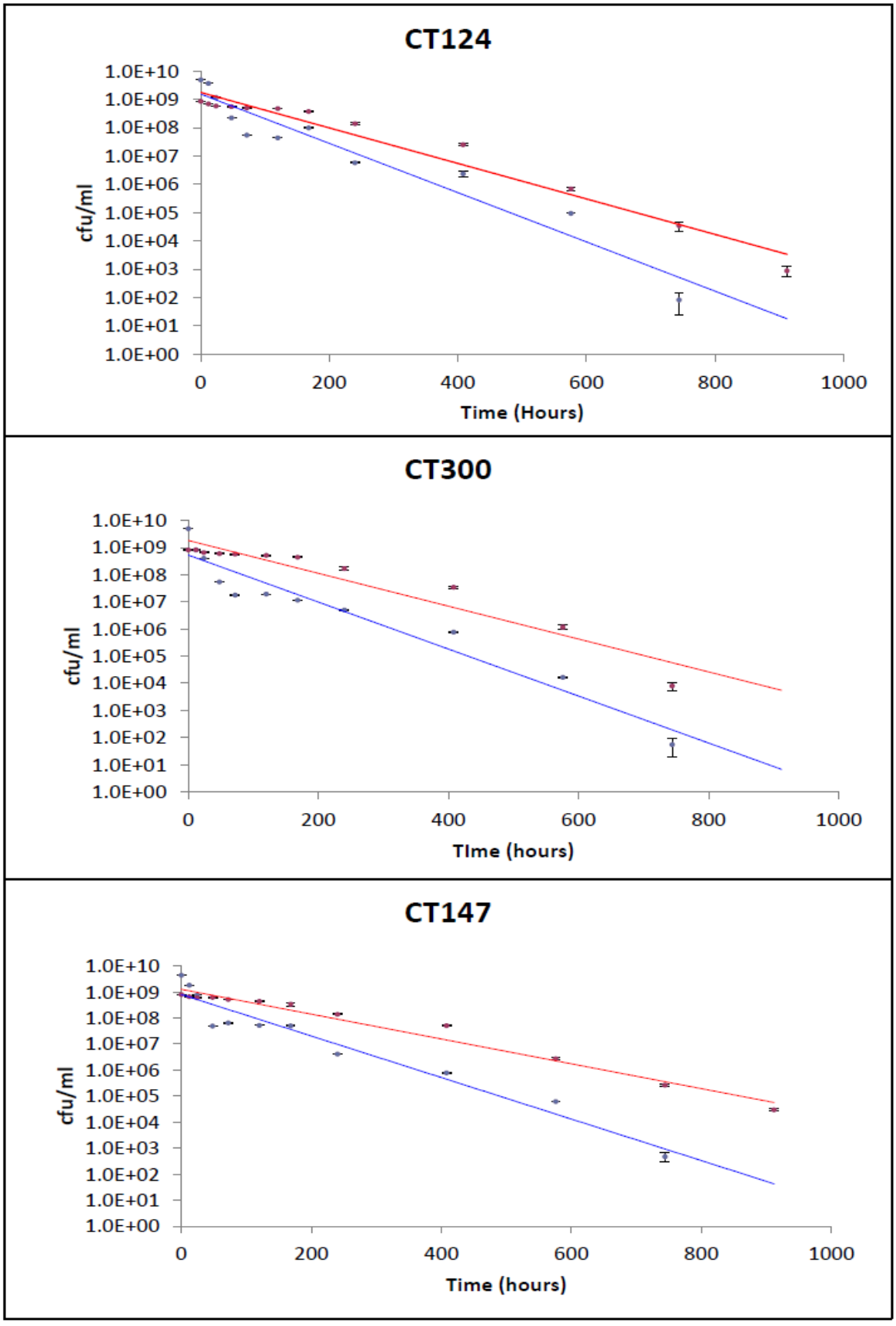

Figure 28a: Linear regression of water (blue) and BSA (red) for each isolates in the desiccation assay (continued in pages 149 to 152)

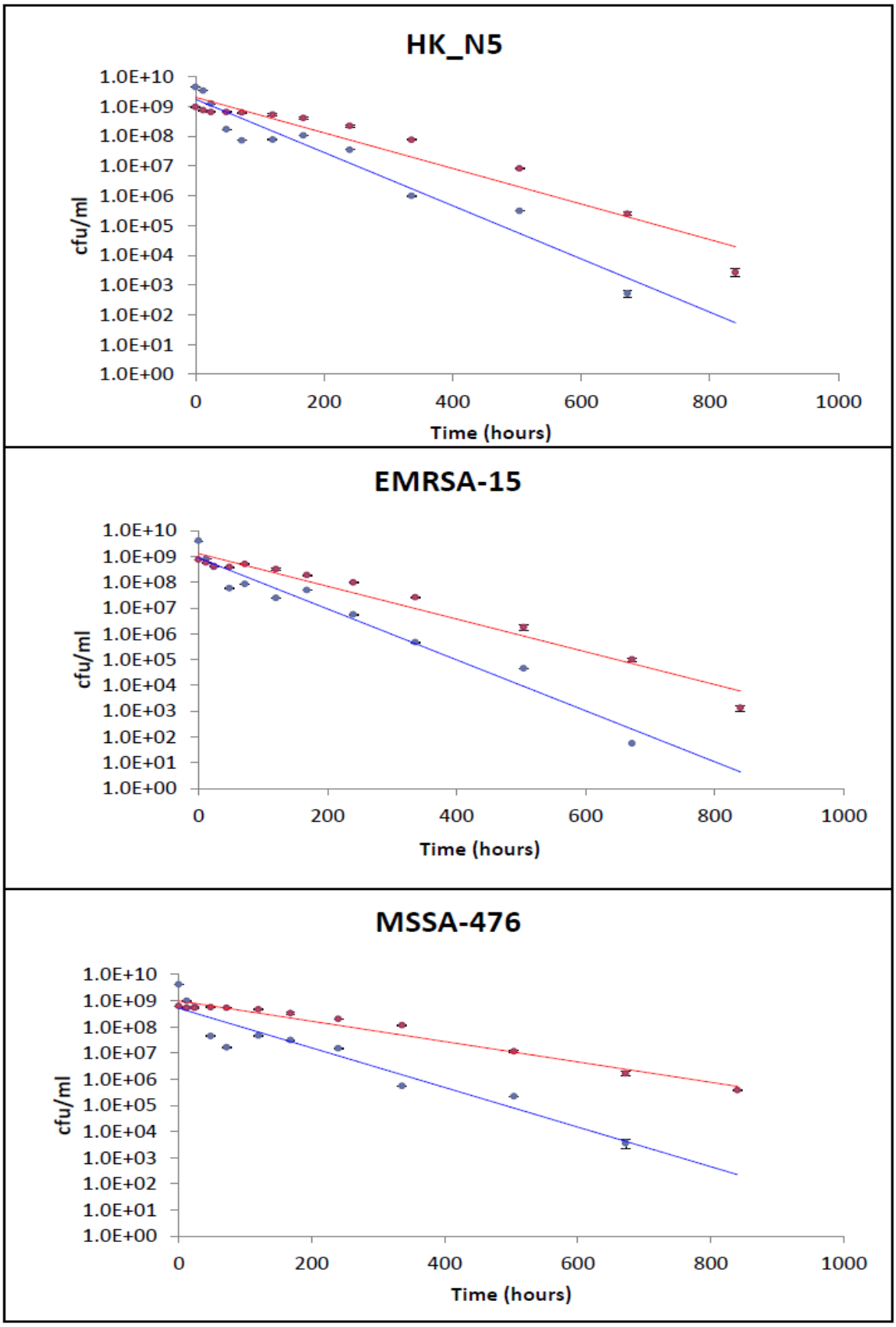

Figure28b: Linear regression of water (blue) and BSA (red) for each isolates in the desiccation assay continued.

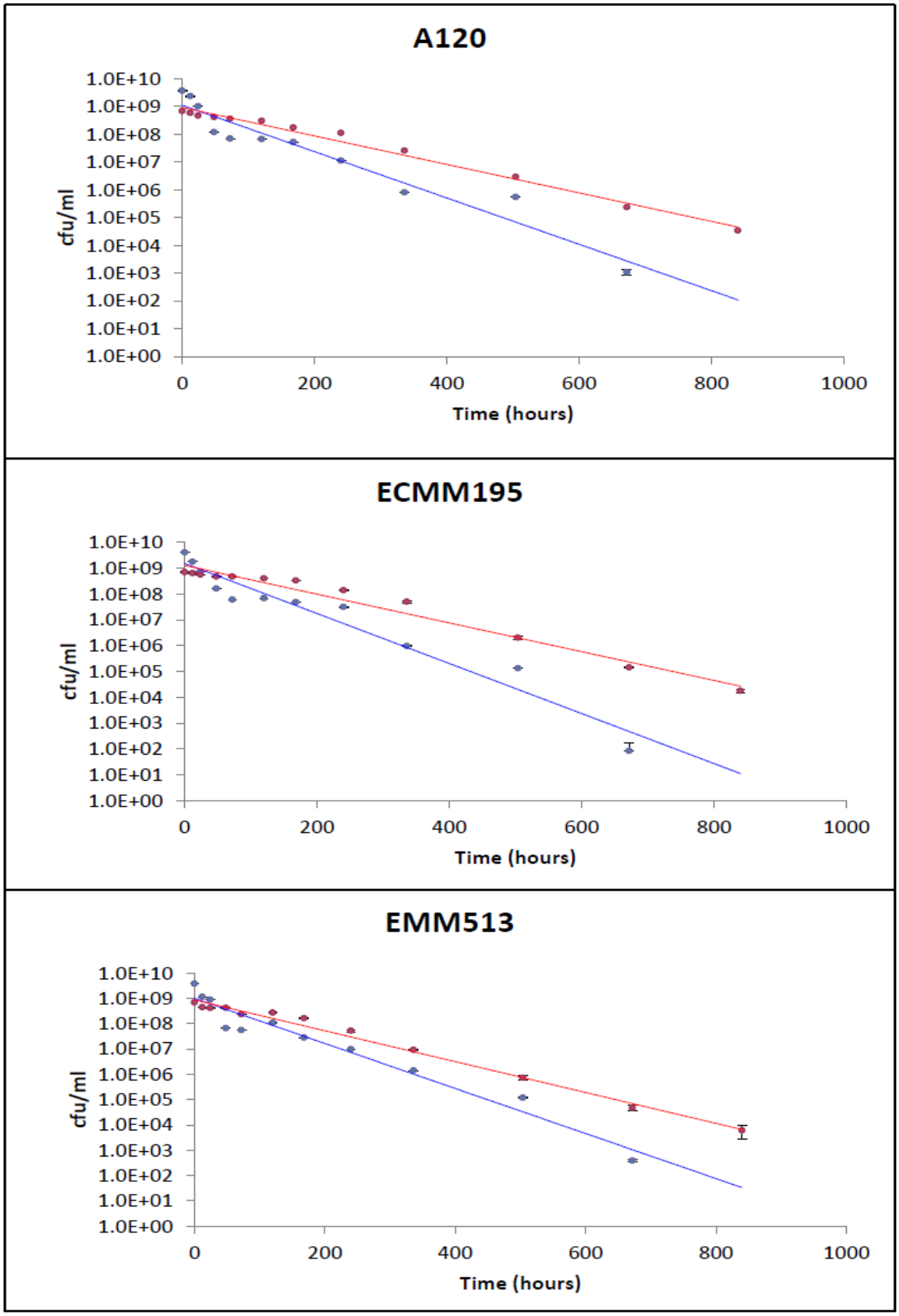

Figure 28c: Linear regression of water (blue) and BSA (red) for each isolates in the desiccation assay continued.

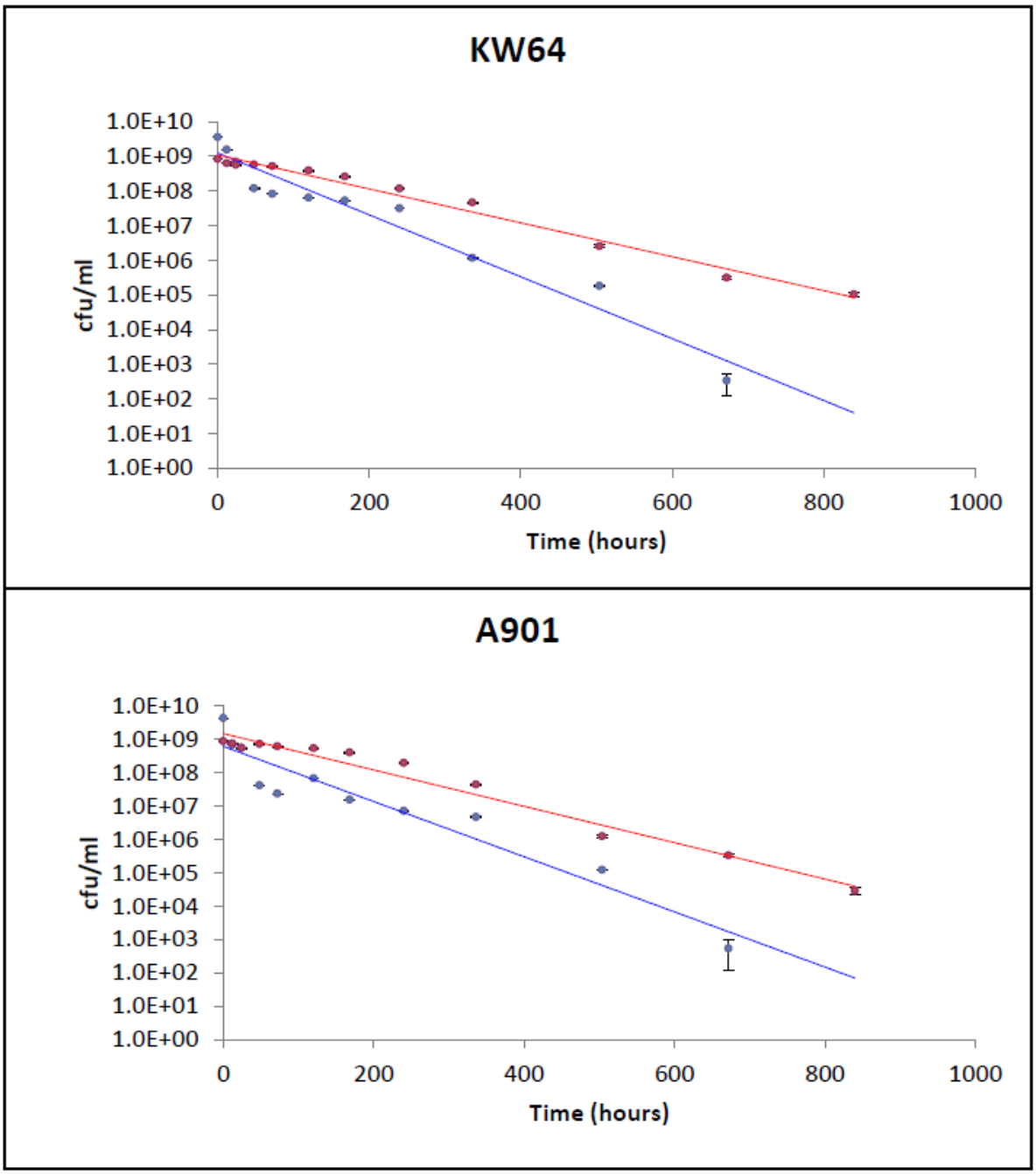

Figure 28d: Linear regression of water (blue) and BSA (red) for each isolates in the desiccation assay continued.

Isolate	Regression line equation	
	Water	BSA
CT124	$y = 2E+09e^{-0.02x}$	$y = 2E+09e^{-0.014x}$
CT300	$y = 5E+08e^{-0.02x}$	$y = 2E+09e^{-0.014x}$
CT147	$y = 8E+08e^{-0.018x}$	$y = 1E+09e^{-0.011x}$
HK_N5	$y = 2E+09e^{-0.018x}$	$y = 2E+09e^{-0.012x}$
E15-C	$y = 7E+08e^{-0.02x}$	$y = 1E+09e^{-0.013x}$
MSSA-476	$y = 5E+08e^{-0.015x}$	$y = 9E+08e^{-0.008x}$
A120	$y = 9E+08e^{-0.017x}$	$y = 9E+08e^{-0.011x}$
ECMM195	$y = 1E+09e^{-0.02x}$	$y = 1E+09e^{-0.011x}$
EMM513	$y = 8E+08e^{-0.018x}$	$y = 8E+08e^{-0.013x}$
KW-64	$y = 1E+09e^{-0.018x}$	$y = 1E+09e^{-0.01x}$
A901	$y = 5E+08e^{-0.017x}$	$y = 1E+09e^{-0.011x}$

Table 24: Slope rate of water and BSA for each isolate in the desiccation assay in the form y = c +b*x.

3.6.3.1 WGS linked analysis

SNPs occurring in the core genome of each of the isolates in the assay were scrutinized for any possible markers for changes in desiccation tolerance. As detailed previously, all of the BHM clade isolates had a SNP in the *grlB* gene however this did not appear to cause any notable changes in desiccation tolerance. No other SNPs were detected in genes associated with desiccation tolerance or stress resistance with the exception of isolate CT300 which demonstrated a non-synonymous SNP in the *crtN* gene that comprises part of the staphyloxanthin operon. This guanine to adenine (G to A) SNP at position 2661949 caused a codon change from GCG to ACG resulting in an amino acid change from alanine to threonine. This amino acid change appears to have ablated staphyloxanthin activity and is supported phenotypically, as CT300 has pale white colonies when grown on ISO sensitest and CBA blood agar (compared to the classical golden yellow of *S. aureus*).

4 Discussion

4.1 Discussion Introduction

The follow section outlines the introduction of the thesis discussion, highlighting areas where this research has investigated novel areas.

MRSA has a highly clonal population structure consisting of distinct sequence types that are endemic or epidemic across different regions of the world. One such lineage is ST22 or EMRSA-15; a HA-MRSA clone that is dominant in the UK and is currently expanding across Europe, Australia (Coombs *et al.*, 2011, Coombs *et al.*, 2014) and into the Middle East (Udo *et al.*, 2014, Udo *et al.*, 2016). As EMRSA-15 continues to expand it is important to investigate both its evolutionary history and continuing epidemiology and evolution to understand how it has evolved in the past and how this will influence its continuing spread across the globe.

Previous studies have used WGS phylogenetic analyses to investigate the emergence and evolution of EMRSA-15 (Holden et al., 2013) and the national epidemiology of EMRSA-15 across the UK (Holden *et al.*, 2013, Reuter *et al.*, 2016). To date no study has extensively investigated the phylogeny of EMRSA-15 in the region it is proposed to have emerged in (Birmingham and the West Midlands). Here I have collected and whole genome sequenced 188 clinical and screening EMRSA-15 isolates from three Birmingham hospitals between c1985 and 2015 to give an intensive, long-time spanning snapshot of the Birmingham EMRSA-15 population. This has not only expanded on previous study time frames but has included a single ciprofloxacin

sensitive ST22 isolates that was sampled from a West Midlands hospital during the mid 1980s and is the first study to include any ST22 isolate from this time period.

The international spread of EMRSA-15 has also been demonstrated in WGS studies. EMRSA-15 has been exported from the UK to Germany, Portugal, the Czech Republic and Singapore followed by clonal expansion within these countries (Holden *et al.*, 2013, Aanensen *et al.*, 2016, Hsu *et al.*, 2015).

By combining Birmingham sequenced isolates with 23 and 14 sequenced isolates from Guernsey and Hong Kong respectively (the first inclusion of isolates from these two location in a phylogenetic analysis) alongside 227 publically available UK and international genomes from the Holden and Reuter papers (Holden *et al.*, 2013) (Reuter *et al.*, 2016) I have generated a time linked MCC phylogeny of 437 EMRSA-15 genomes. At the time of writing this is the largest number of EMRSA-15 isolates used to generate a MCC tree and is the longest time spanning Bayesian phylogenetic study of UK and internationally sourced EMRSA-15 isolates (1985 to 2015).

Evidence for the clinical use of fluoroquinolones and the evolution of MRSA has been detailed in previous literature. Holden and colleagues have suggested clinical trials of ciprofloxacin in the West Midlands led to the emergence of the ciprofloxacin resistant EMRSA-15 lineage (Holden *et al.*, 2013) while studies examining MRSA bacteraemia rate and fluoroquinolone usage have shown correlation between increasing fluoroquinolone use and increasing MRSA infection rate (Charbonneau *et al.*, 2006, Parienti *et al.*, 2011).

The data presented here is the first to extensively interlink fluoroquinolone prescription data with a phylogenetic reconstruction of EMRSA-15 over time.. This has

allowed a comparison of how fluoroquinolone usage and phylogenetic diversity in a single hospital has changed over a 25 year period identifying changes in flouoroquinolone use that may affect the phylogenetic structure of EMRSA-15 in Birmingham.

In addition to fluoroquinolone resistance, EMRSA-15 has phenotypic characteristics that make it a well adapted hospital pathogen. MRSA has been shown to survive in the hospital environment on specific surfaces such as bed rails, linen or medical equipment, therefore providing an environmental reservoir for infection. As such, increased tolerance to desiccation has important implications for EMRSA-15 in the clinical setting, where the bacteria are exposed to harsh environmental conditions that cause drying or desiccation, and plays a key role in MRSA persistence in the hospital environment (Chemaly *et al.*, 2014). To date no study has examined the desiccation tolerance of sub strains of EMRSA-15 with differing ciprofloxacin resistance or strains sampled from different geological and climatological niches. Here EMRSA-15 sub lineages have been identified from our phylogenetic analysis along with EMRSA-15 isolates sampled from Hong Kong and Kuwait to compare their desiccation tolerance to each other in a controlled laboratory environment.

4.2 Pipeline validation

As WGS technology has expanded the scope of bioinformatics programmes available to analyse the produced data has expanded with it. This has led to a myriad of publicly available programmes and pipelines, Each of these programmes have unique features

but ultimately perform similar functions. Unlike with other scientific disciplines, public forums play an important role in the review and development of programmes as well as how to implement them along with any caveats or faults. When used in conjunction with published literature this can provide a high degree of review/confidence in the selected programme.

This study has constructed a unique pipeline to trim then map Illumina paired end reads to the EMRSA-15 reference genome HO 5096 0412 (Koser *et al.*, 2012b), determine SNPs in the core genome of each isolate and remove recombinant SNPs to generate an aligned fasta file for application to BEAST. When compared to four WGS studies investigating EMRSA-15 epidemiology the pipeline produced here was unique but found to employ similar protocols for phylogenetic tree building and the use of BEAST (Table 16)(Holden *et al.*, 2013, Reuter *et al.*, 2016, Hsu *et al.*, 2015, Aanensen *et al.*, 2016).

All pipelines, including the one described here mapped trimmed reads to the reference genome HO 5096 0412, that was part of a UK neonatal ward outbreak and was published in 2012 (Koser *et al.*, 2012b). This is due to it being the only the fully annotated EMRSA-15 reference genome available and therefore the only suitable genome for SNP analysis of EMRSA-15 genomes.

Each study showed a range of average read coverages produced from their respective sequencing platforms, however debate arises around what level of coverage is required for WGS analyses. Appropriate coverage depth of sequenced reads is dependent on the analysis being performed and mapping based approaches require less coverage than de-novo assembly. Most coverage considerations come from mammalian genomic studies, 15x for homozygous SNP determination and 33x for heterozygous SNP

determination, as they were the first to investigate NGS SNP typing applications (Wang *et al.*, 2008, Ahn *et al.*, 2009). These values have been lower in subsequent years as sequencer read length has increased, therefore reducing the need for higher coverage depth. Furthermore eukaryote genomics require higher coverage due to multiple chromosomes and diploid genomes, with coverage of single choromosomal bacteria being downscaled (Ekblom and Wolf, 2014, Sims *et al.*, 2014). Studies with eukaryotic genomes have also demonstrated a trade of between increasing sample size and reduced sequence coverage, where the error rate from low read coverage is mitigated from data provided by the other samples in the population (Le and Durbin, 2011) (Li *et al.*, 2011).

The average coverage in this study was 60.10x which is comparable to other MRSA (Reuter *et al.*, 2016) (Driebe *et al.*, 2015) or bacterial studies (Koser *et al.*, 2014, Cairns *et al.*, 2015). Here we have used Qualimap with a 1000bp sliding window to assess coverage of each isolates across the HO 5096 0412 reference. This has given more accurate sequence coverage values for each isolates and ensured complete coverage across the reference genome for give accurate and reliable SNP determination.

Only SNPs from the core genome, the ubiquitous set of genes that are present in all genomes of the study, were used to build evolutionary phylogenetic trees with SNPs arising in MGEs (including the SCCmec element, prophages, plasmids and pathogenicity islands (Lindsay *et al.*, 2012)) and repetitive regions being excluded. This method has been applied in all EMRSA-15 phylogenetic studies (Holden et al., 2013, Reuter et al., 2016, Hsu et al., 2015, Aanensen et al., 2016) due to the need to build phylogenies on reliable SNPs and exclude those in MGEs and repetitive regions which may give misleading analyses. Therefore the core genome provides the most consistent set of

SNPs for evolutionary reconstruction (Planet *et al.*, 2016), but can lose additional phylogenetic information that would be present in the MGE genes.

4.2.1 Homologous recombination

The impact of homologous recombination on the evolution of bacteria has been well documented (Feil and Spratt, 2001) and can effect the accuracy of phylogenetic reconstructions due to a single recombination event being able to introduce multiple polymorphisms, therefore confounding the ability to accurately measure the rate of nucleotide substitutions arising by point mutation. Furthermore recombination occurring between two isolates in a dataset may make them appear to be more closely related than they actually are and can reduce the accuracy of phylogenetic signal (Croucher *et al.*, 2013, Posada and Crandall, 2002). Therefore it is prudent to identify and remove recombinant SNPs from phylogenetic analyses. This study used the Gubbins tool (Croucher *et al.*, 2015) to identify blocks of recombination arising in each of the 437 genomes. Gubbins is similar to other recombination determining programmes such as ClonalFrame (Didelot *et al.*, 2010) or BRATNextGen (Marttinen *et al.*, 2011) which identify regions of high density SNPs occurring in the genome. Gubbins has the advantage however of converging on results significantly faster than other programmes (Croucher *et al.*, 2015) and is tailored for use with clonal populations. A drawback of using SNP density methodology for recombination determination is that SNPs naturally arising in high-density regions or those occurring from sequencing error cannot be distinguished from those arising by recombination, allowing a margin of error in recombinant SNP determination. This is likely to be minimal in this dataset due to the

large sampling population and the low error rate reported for Gubbins (Croucher *et al.*, 2015).

The number of recombinant SNPs in my dataset was low (3.15% of determined SNPs). This value is similar to those observed in other EMRSA-15 studies (Holden *et al.*, 2013) but is higher than in ST239, which has been seen to range from 0.88% (Harris *et al.*, 2010) to 1.1% of core genome SNPs (Hsu *et al.*, 2015). This could be due to ST239 and ST22 being endemic to different regions of the globe and are therefore under different evolutionary selection pressures (climate environment, hospital environment, antibiotic regimes). Another possible reason is the sampling population examined in each study. Inclusion of a greater number of diverse isolates will give more representation across lineages compared to studies that have included a large number of highly related or virulent isolates (Didelot and Maiden, 2010). This study presented here has both included a larger number of isolates and both clinical and screening isolates and therefore may give a more accurate representation of recombination. While specific rates between MRSA lineages may vary, low rates of recombination are consistently reported in MRSA. This indicates recombination within MRSA lineages is detectable but occurs infrequently, demonstrating the highly clonal nature of MRSA sequence types (Planet *et al.*, 2016, Kuhn *et al.*, 2006).

4.2.2 Bayesian phylogenetics

Several methods of phylogenetic tree construction are available, all of which display data in different ways. One of these methods is to use Bayesian analysis with Marcov Chain Monte Carlo (MCMC) statistical modeling with the Bayesian Evolutionary Analysis

Sampling (BEAST) platform to create maximum clade credibility (MCC) trees. Bayesian methods are becoming increasingly used due to the availability of large SNP datasets to infer accurate timelines of bacterial evolution. Grey *et al* (2011) were the first authors to use Bayesian inference to understand MRSA epidemiology by applying publically available SNP data from the Harris *et al* 2010 global ST239 study to the BEAST platform (Gray *et al.*, 2011) to demonstrate the temporal evolution and global transmission of ST239. This method is considerably more complex than more standard methodologies such as ML or NJ trees (Huelsenbeck *et al.*, 2001), but has the added value that it allows the temporal divergence of isolates to be determined, therefore providing a greater understanding of how a species or strain has evolved over time. This is particularly appropriate for the current study outlined here, where the aim was to investigate the evolution of a single lineage over time and determine when the strain emerged. This information would not be able to be gained from ML or NJ phylogenetic techniques.

Before data is applied to BEAST it is important to evaluate if a robust ML phylogeny has been created. Bootstrapping remains one of the most common methods for demonstrating the robustness of a phylogenetic tree since it was first applied to phylogenetics in 1985 (Felsenstein, 1985) and is a measure of how often the tree topology is recreated when randomly generated many times over (ideally 500 times or above). The tree generated in this study had high bootstrap values (mode 100) which indicated a robust phylogeny had been generated from the WGS data before it was applied to BEAST.

BEAST is a complex programme to operate and it has several caveats that need to be considered. Firstly, the extensive choice of parameters, models and priors require

significant understanding to be implemented appropriately due to accurate inference relying on suitable model selection and calibrations (along with informative sequence data). The selection of an appropriate molecular clock for the dataset is critical and involves determining a balance between the complexity of the model and the fit of the dataset to the model. This can lead to a "trial and error" approach to generate the desired outcome and variation between different studies but can also result in misleading divergence time estimates if a poor model is chosen. Due to the exact calibrations of BEAST not being defined in published literature direct comparisons in BEAST methodology are not possible however it should be noted that differences in molecular clock or population models can infer differences in the MCC tree.

A variety of molecular clocks have become available as molecular evolutionary studies have expanded over the past 20 years. The most basic clock is the strict molecular clock (or global molecular clock) which assumes that the rate of evolution is the same for every branch of the tree. In contrast the relaxed molecular clock assumes heterogeneity in the rate along branches of the tree and allows different substitution rates on each branch (Ho and Duchêne, 2014). Gray and colleagues (Gray *et al.*, 2011) used a combination of clock and population models to demonstrate how different model combinations can give different tMRCA values (ranging 21 years) and the superiority of the relaxed molecular clock compared to the strict molecular clock for tMRCA determination.

The Path-O-Gen programme provided by BEAST can analyse ML trees to determine if a strict molecular clock would be appropriate for the dataset by measuring the root to tip distance of each isolate against time. The dataset of 437 genomes applied to Path-O-Gen in this study showed an increasing accumulation of sequence variation over the 25 year

sampling timeframe with only 26 isolates (5.9%) displaying marked variation away from the linear regression line. This small number of outliers is indicative of a clonal population of isolates that have evolved at a similar rate, demonstrating linear specific evolution and that a relaxed clock model should be applied in BEAST.

Statistical methods can also be employed to determine the most appropriate model for a dataset, the most common of which is Bayes factor test (BFT) (Drummond and Rambaut, 2007, Jeffreys, 1935, Kass and Raftery, 1995). BFT calculates marginal likelihoods of each model therefore allowing a pairwise comparison of models and selection. However this in itself has the caveat of the model with the best fit to the dataset may not be the best fit for the biology of the organism being studied. BFT was used here to determine that a relaxed clock model with a constant population model was the most appropriate model for my dataset. This was confirmed with high ESS values from trace files and a hairy caterpillar trace that indicated a robust model had been chosen for my analysis.

Another caveat of implementing BEAST is it is highly computationally demanding and can take weeks to months to complete a 10,000,000 chain run depending on the size of the dataset. Furthermore each genome sequence added to the analysis increasing the time for completion. Many laboratories may not have the computational capabilities to accommodate this analysis, however the increasing demand for bioinformatics has led to several freely accessible cloud based infrastructures to be developed (http://www.climb.ac.uk) which can help alleviate this problem. Due to the high number of genomes applied to the BEAST platform in this study, analysis time for each BEAST

run was approximately 2 months, however the use of cloud based infrastructure allowed these runs to be completed in parallel, drastically reducing analysis time.

For accurate BEAST divergence dating a good representation of isolates across time is required. This is due to reduced accurate where there are large gaps in the timeframe of isolate sampling. For example a period of 10 years with no sampled isolates means that there is no molecular data to estimate evolutionary change over this 10 year time period. By including a large number of isolates over a lengthy timespan this inaccuracy has been reduced and strengthened the phylogenetic reconstruction. Unfortunately no studies, including this one, have many EMRSA-15 isolates sampled between 1980 and 2001 included in analyses. This is because EMRSA-15 isolates from this time period being extremely difficult to source due to two main reasons. Firstly isolates from the 1980s would need to have been stored for over 30 years and collections may have been lost due to poor archiving and maintenance of collections. Secondly the MRSA infection rate was relatively low in the UK until the mid 1990s (Figure 1) and was not considered a public health threat until the late 1990s and early 2000s (Duerden *et al.*, 2015). Indeed as the number of MRSA bacteraemias began to rise in the 1990s and governments around the world introduced national policies to tackle the threat of MRSA, isolates were stored in much greater number for reference and research purposes which is demonstrated by the large availability of MRSA isolates from 2001 onwards in this study and other MRSA evolutionary studies (Reuter *et al.*, 2016) (Holden *et al.*, 2013).

4.3 Global phylogeny

The public health threat of the continued expansion of EMRSA-15 has been widely acknowledged and as such there is a need to investigate the epidemiology of the global EMRSA-15 population. A previous study by Holden and colleagues whole genome sequenced 193 EMRSA-15 isolates sampled between 1991 and 2008 and employed Bayesian phylogenetics and phylogeographics (determining spatial and temporal patterns of population structure using genetic variation (Avise *et al.*, 1987)) to suggest EMRSA-15 evolved in the West Midlands (UK) in the mid 1980s after acquiring ciprofloxacin resistance.

The data presented here has expanded on this knowledge base by analyzing 437 EMRSA-15 genomes that included over 200 isolates from Birmingham and a single ST22 isolate from the region that pre-dates the emergence of EMRSA-15. To date this is the largest WGS study of EMRSA-15 that has longitudinally investigated a single region and contextualized in a national and international setting.

The selection of isolates for inclusion in a phylogenetic study is arguably one of the most important facets of the study due to impact of sample biasing (Planet *et al.*, 2016, Pearson *et al.*, 2009). Unlike previous EMRSA-15 evolutionary studies that have used only clinical isolate collections (Reuter *et al.*, 2016) or collections of isolates that may be biased towards outbreak sampling (Holden *et al.*, 2013), the collection analysed here included 152 screening isolates (34.7% of total genomes analysed). By including a large number of screening isolates alongside clinical and internationally sourced isolates I have generated a more representative cross sectional understanding of the EMRSA-15 population over this time period and reduce the biasing that may be introduced by analyzing only isolates that have caused clinical infection.

4.3.1 Evolutionary rate

Investigating the rate of evolution of bacteria is important to be able to understand how isolates have evolved over time. By knowing the rate at which a strain has evolved one can infer the number of SNPs that have arisen between isolates since they shared a common ancestor. By using BEAST I have been able to determine the evolutionary rate of the EMRSA-15 population analysed in this study to be 1.38 $x10^6$ substitutions per site per year or the equivalent to approximately 5 mutations across the genome per year. This value is within the 95% confidence intervals described by Holden et al in 2013 (Holden *et al.*, 2013) who demonstrated a rate of 1.30 $x10^6$ substitutions per site per year and Hsu *et al* in 2015 who demonstrated a rate of 1.4 $x10^6$ substitutions per site per year from a Singaporean collection of EMRSA-15 (Hsu *et al.*, 2015). The inclusion of a large number of the Holden et al 2013 isolates in this analysis may bias the rate result, however the BEAST analysis that only included Birmingham sampled isolates had a similar rate of 1.36 $x10^6$. This concordant collection of evidence would indicate that EMRSA-15 has a steady evolutionary rate across different geographical areas, Birmingham versus Singapore for example, but also that it appears to evolve slower than other MRSA lineages. ST239 has been reported to evolve at a rate of 2.2 $x10^6$ to 3.3 $x10^6$ substitutions per site per year ((Harris *et al.*, 2010, Hsu *et al.*, 2015, Castillo-Ramírez *et al.*, 2012)) and EMRSA-16 has been reported at 1.42 × 10^6 substitutions per site per year (McAdam *et al.*, 2012) which is more similar to the rate of EMRSA-15. The reasons for this potential difference in evolutionary rate between lineages is similar to the reasons described for homologous recombination (section 4.2.1). Different methodologies have different accuracies, for example Bayesian inference with BEAST (Hsu *et al.*, 2015, Holden *et al.*, 2013, McAdam *et al.*, 2012) is more accurate than root to

tip and sampling date linear regression analysis (Harris *et al.*, 2010) (Young *et al.*, 2012). The isolates included in a study will also influence rate determination. Evolutionary rate is not fixed in time and space, with environment specific selection pressures, such as colonization versus blood stream infection, or evolutionary bottlenecks changing the mutation rate. Therefore determining mutation rate from isolates that have only been sampled from a single sub-lineage, outbreak or clinical infections may give different results compared to those from a more varied sampling source collection. Here I have included clinical, screening and globally sourced EMRSA-15 isolates and have a greater total of isolates analysed, therefore may give a more accurate rate of the EMRSA-15 population by reducing sample bias.

4.3.2 The evolution of ciprofloxacin resistance and the EMRSA-15 lineage

The historical ciprofloxacin sensitive isolates from CHB sampled in c1985 was not only related to the other UK ciprofloxacin sensitive genomes, due to it's basal positioning on the tree, but was also highly related to the first reported historical EMRSA-15 genomes sampled in 1991 to 1993 from Birmingham. These historical Birmingham EMRSA-15 isolates were 50 SNPs from the CHB historical isolate which is consistent with the SNP accumulation rate discussed in section 4.3.1 over a six year time period. Furthermore the historical CHB progenitor isolate was sampled in a hospital that carried out the first ciprofloxacin trials in the West Midlands (Crump *et al.*, 1983) indicating the ST22 progenitor population was in present in a hospital where ciprofloxacin was being trialed. This combination of topology and SNP distance is indicative that a hospital associated ciprofloxacin sensitive ST22 population (lacking the

Ser80Phe mutation in *grlA* and Ser84Leu in *gyrA* mutation) that was highly related to the emergent EMRSA-15 was circulating in the West Midlands. This would therefore indicate that ciprofloxacin resistance did emerge in the West Midlands rather than be imported from elsewhere in the UK (and subsequently transmitted onward from Birmingham).

To date this is the first report of the whole genome sequencing and phylogenetic analysis of a hospital sampled ciprofloxacin sensitive ST22 isolate that predated 1991. It is therefore a key isolate in the determination of EMRSA-15 evolution by providing genetic information from the time EMRSA-15 emerged and allows a more robust phylogenetic trees to be built. The addition of more ciprofloxacin sensitive or resistant ST22 isolates sampled before 1991, ideally from other UK or international locales, to this phylogeny would further enhance my findings by giving more precise estimates of the time EMRSA-15 evolved. As stated in section 4.2.2 isolates from this time period are rare and no isolates from elsewhere in the UK from this time could be sourced in either this study or the study by Holden and colleagues (Holden *et al.*, 2013). Therefore, theories of simultaneous evolution of EMRSA-15 in other UK locales or the importation of the ST22 progenitor population from elsewhere in the UK before ciprofloxacin trials began cannot be excluded.

The emergence of ciprofloxacin resistant ST22 (EMRSA-15) was dated as approximately 1984 in this study. This date is two years earlier than was proposed by Holden *et al* however 95% HPD intervals for both this divergence date and the Holden and colleagues date overlap, indicating relative concordance. This difference in divergence dating may be accounted for by the increased number of isolates in my phylogeny and

the inclusion of the ciprofloxacin sensitive ST22 isolate from Birmingham providing increased time divergence dating accuracy at this time point on the MCC tree.

The possible earlier evolution of ciprofloxacin resistance in ST22 has an impact on current theories of EMRSA-15 evolution. Holden and colleagues propose that ciprofloxacin trials occurring in the West Midlands (UK) region may have been the selection pressure for the evolution of ciprofloxacin resistance based on three published ciprofloxacin trials in 1983 (Crump *et al.*, 1983) and 1986 (Silverman *et al.*, 1986) (Finch *et al.*, 1986) coinciding with a ciprofloxacin resistance emergence date of 1986. Only the trial by Crump *et al* in 1983 pre-dates my EMRSA-15 emergence time point of 1984 and was a trial conducted on only six patients at CHB (then known as Dudley Road Hospital) but interestingly occurred in the hospital the EMRSA-15 progenitor isolate was sampled from. It is possible that ciprofloxacin use was limited in clinical trials before 1986 and a resistant phenotype could have emerged in 1984 but did not spread until fluoroquinolone use became more frequent in clinical trials in 1986 that included larger numbers of patients (26 patients (Silverman *et al.*, 1986) and 52 patients (Finch *et al.*, 1986)). Furthermore significant early branching of the MCC phylogenetic tree is evident from circa 1985 to 1986 and may be accounted for by the increasing use of pre-licenced ciprofloxacin. It should however be noted that time divergence dating uncertainty increases the further back in time the phylogeny goes and neither the data presented here or in the Holden *et al* 2013 paper are exact dates for the emergence of EMRSA-15.

4.4 Global phylogenetic structure

EMRSA-15 continues to spread internationally and is currently the fastest growing HA-MRSA lineage in Europe (Grundmann *et al.*, 2014) and has become a healthcare threat in the Middle East (Udo *et al.*, 2016, Udo *et al.*, 2006, Goudarzi *et al.*, 2016) and Australasia (Hsu *et al.*, 2007b, Coombs *et al.*, 2014, Groves *et al.*, 2016, Hsu *et al.*, 2005).

The inclusion of EMRSA-15 genomes from new international sources being analysed alongside previously WGS phylogenies is required to monitor the continuing expansion of the strain. By including the isolates from Hong Kong, Guernsey and the UK with genomes from ten additional countries I have created the largest global phylogeny of EMRSA-15 to date.

The global phylogenetic reconstruction presented here is highly concordant with the Holden *et al* paper (Holden *et al.*, 2013), where the international genomes were sourced, and demonstrates genomes from specific countries clustering distinctly together indicating importation and clonal expansion. Three recent whole genome studies that used ML (Aanensen *et al.*, 2016) and MCC (Holden *et al.*, 2013, Hsu *et al.*, 2015) phylogenetic trees have demonstrated exportation of EMRSA-15 from the UK followed by clonal expansion to become the dominant clone within healthcare systems that previously had no significant ST22 MRSA burden. This is corroborated in my analysis where UK isolates that were less than 40 SNP distant were directly basal to each cluster of German, Portugese and Singaporean isolates in the MCC phylogeny. In contrast to Aanensen and colleagues a Guernsey isolate was basal to the Portuguese cluster (36 SNP distance) instead of a UK mainland isolate (over 100 SNPs (Aanensen *et al.*, 2016). This could potentially suggest Guernsey as an exportation source to Portugal rather than

directly from the UK, however the sampling bias of limited number of isolates from both Guernsey and Portugal in this analysis means complex routes of importation of EMRSA-15 to Portugal may have occurred.

Both the Guernsey and Hong Kong isolates demonstrated geographically restricted clustering (Hong Kong isolates are discussed in the section below). Guernsey is a small British island dominion in the English Channel where EMRSA-15 is also the dominant MRSA clone. The phylogenetic structuring of EMRSA-15 in Guernsey showed a localized EMRSA-15 population predominating alongside EMRSA-15 isolates that were more related to those of the UK mainland. This echoes the findings of Birmingham, where a distinct Guernsey clone is circulating alongside a more genetically diverse population of EMRSA-15 that are related to other UK circulating strains. Movement of patients and healthcare workers between Guernsey and the UK mainland is common. Therefore the distinct Guernsey clone may have been imported from the UK and then clonally expanded in the Guernsey healthcare system at a similar time to when EMRSA-15 was expanding across the rest of the UK (c1996 from this analysis). Additional sporadic introductions of EMRSA-15 from the UK may also account for the UK related clones in the population. Some biasing is however present in the Guernsey collection due to seven of the ten isolates in the Guernsey branch of the phylogenetic tree being from an ITU outbreak. However three non ITU isolates also clustered to this branch suggesting it is a real phenomenon.

4.4.1 Transmission of EMRSA-15 in East Asia

The 14 isolates from Hong Kong showed evidence of region specific clustering following a single importation event. Unlike other countries the importation of EMRSA-15 to Hong Kong appears to be from Singapore and not directly from the UK. Unlike in the UK, ST239 is the dominant MRSA lineage across Asia (Ko *et al.*, 2005, Feil *et al.*, 2008, Chen and Huang, 2014) including Hong Kong and formerly in Singapore up to the mid 2000s. Between 2003 and 2015 there was a notable increase in ST22 MRSA strains being isolated from clinical specimens in Singapore where EMRSA-15 in now the dominant lineage (Hsu *et al.*, 2005, Hsu *et al.*, 2015, Hsu *et al.*, 2007b). Using BEAST on a collection of 87 isolates from four Singaporean hospitals, Hsu and colleagues demonstrated that this replacement of ST239 with ST22 likely occurred from an importation event from the UK in 2001, which is supported by my analysis. The BEAST analysis presented here indicates that approximately one year after this importation to the Singaporean healthcare system there was an exportation event to Hong Kong. This transmission event is strengthened by evidence of only two SNPs differentiating the Singaporean and Hong Kong clusters of the MCC tree.

This is the first report of UK, Singapore and Hong Kong EMRSA-15 isolates being analysed together and highlights how movement of strains between healthcare systems can occur across large geographical areas. A similar phenomenon has been demonstrated with ST239 being imported to London (UK) where it caused an outbreak of invasive infections in ITU (Edgeworth *et al.*, 2007) but failed to become established in the healthcare system. This exportation was subsequently found to have likely been imported from Singapore using a BEAST analysis (Hsu *et al.*, 2015). Unlike in Singapore EMRSA-15 has not become established as the dominant lineage in Hong Kong (where

ST239 remains the dominant clone) despite having been introduced to the respective healthcare systems at a similar time (2001 and 2002 [Figure 10]). Due to the high degree of genetic similarity between the Hong Kong and Singaporean isolates it is prudent to suggest differing healthcare practices between Singapore and Hong Kong rather than genetic mutation and adaptation are the reason for the poor success of EMRSA-15 in Hong Kong. Regional differences in ST239 *spa* type have been demonstrated in China (Chen *et al.*, 2014) with isolates from Hong Kong being more related to "Asian clade" ST239 *spa* type t037 compared to those from Beijing which are more related to "Turkish clade" ST239 *spa* type t030. Shang and colleagues have shown that *spa* type t037 grew slower and had a reduced viable count in a murine kidney infection model compared to t030 (Shang *et al.*, 2015) showing differences in infectivity between the ST239 *spa* types. Therefore it is possible the ST239 population in Hong Kong is not as easily outcompeted in the healthcare niche as those in Singapore, however no spa tying data is available from Singapore, therefore making comparisons difficult.

Increased sampling from Asian countries and continued public access to WGS data will increase the understanding of MRSA movement across Asia as a whole, where genome studies are largely limited to countries such as Singapore, Japan or China (Kong *et al.*, 2016).

4.4.2 Birmingham Phylogeny

Analysis of the global and Birmingham phylogenetic trees has demonstrated a distinct EMRSA-15 population across the large Birmingham UK conurbation.

The Birmingham EMRSA-15 population structure comprised of a dominant, localized, highly clonal lineage (BHM clade) alongside a more diverse population of isolates that were more related to other UK strains. Reuter *et al* (2016) demonstrated similar geographical structuring of EMRSA-15 in the UK on a national scale. Using an extensive collection of bacteraemia isolates they were able to demonstrate the emergence of region specific sub clones but had only 16 isolates from Birmingham (all from CHB) in their West Midlands sample population (Reuter *et al.*, 2016, Holden *et al.*, 2013). Furthermore BSAC submitting centres may not accurately represent hospitals in each referral region. For example a single hospital centre submitted isolates to BSAC from the "South Central" region and centres submitting from the "North Central" region being localized to the east coast (Newcastle) (Reuter *et al.*, 2016).

Here I have performed a higher resolution analysis by extensively sampling across Birmingham and applied Bayesian inference to generate this phylogeny. Furthermore I have also included the publically available genomes of the West Midlands isolates from the Reuter study to further contextualize the Birmingham EMRSA-15 population in the West Midlands region.

The BHM lineage was highly localized to Birmingham but contained two isolates sampled from Coventry (West Midlands, UK) and a single isolates from Shrewsbury (West Midlands, UK), which may have been exportations from one of the three Birmingham hospitals. As stated above Reuther and colleagues demonstrated a similar phenomenon in their regional sampling framework. Due to the authors not using a

BEAST analysis, which would be difficult to achieve with over 700 isolates, dating the divergence of their regional specific sub-clones its not possible, however it is feasible they may have emerged at a similar time to the BHM clade (c1995) (Reuter *et al.*, 2016).

The BHM clade dominated the Birmingham healthcare system, comprising approximately half of the total Birmingham sampled isolates and constituted over 50% of isolates sampled in 2003, 2005 to 2008 and 2014 to 2015. Bayesian inference estimated the divergence date of the BHM clade to be circa 1995 and less than 10 years after it emerged there appears to have been a cladal shift towards BHM dominance in the Birmingham MRSA population. Due to the lack of isolates sampled before 2001 from Birmingham an importation of the BHM clade from elsewhere in the UK cannot be excluded, however the lack of proximal clustering of any of the UK isolates near the BHM clade would suggest this is unlikely.

A phylogenetic study using ML methodology of 76 ST239 isolates prospectively sampled over three months in a single hospital in Thailand (Tong *et al.*, 2015a) showed that there was significant diversity in the HA-MRSA population over time. Three major ST239 sub-clades were present that were each distinguishable by over 100 SNPs, with no single stable clone being maintained within the hospital. In contrast my findings show that a single, highly clonal EMRSA-15 sub-strain has emerged and been maintained in the Birmingham hospital population for 20 years; the first description of a clone being maintained for such a period of time. This indicates the BHM clone is well adapted to the Birmingham healthcare system and lack of transmission out with the Birmingham conurbation may be due to care in Birmingham is predominantly provided by the three hospitals analysed here with patients being transferred between HEFT,

QEHB, and CHB. Similar patient demographics may also account for the regional clones described by Reuter and colleagues 2016.

Delineation of the sampling hospital of each of the Birmingham isolates showed isolates linked in space and time to cluster, as has frequently been demonstrated in WGS studies (Harris *et al.*, 2013, Koser *et al.*, 2012b). However there was a lack of hospital specific structuring, with isolates from all three hospitals being distributed throughout the tree. Hsu *et al* 2015 demonstrated a similar phenomenon in Singapore across four hospitals in a single healthcare system over a 10 year time period for ST239 and ST22 (Hsu *et al.*, 2015) and studies in the USA using over 300 sterile site infection isolates from four hospitals (Long *et al.*, 2014) and 30 clinical isolates from seven hospitals (Prosperi *et al.*, 2013) have demonstrated similar results. The concordance of this study with these reports that used both Bayesian and ML phylogenetics with multiple hospital sites and long temporal sampling would indicate that transmission between hospitals across restricted geographical areas is frequent and that a shared patient population is present in the region.

4.4.3 Evaluation of WGS and VNTR typing

VNTR typing and WGS examine different regions of the genome, the accessory and core genome respectively, however to date no published study has investigated the relationship between the two techniques in MRSA.

The clonal nature of HA-MRSA lineages means that VNTR typing, and other genotyping techniques, have a reduced ability to discriminate between isolates within a single HA-MRSA lineage. This is demonstrated here with EMRSA-15, where in all 210

isolates no variability in repeat number was observed at five of the seven loci. The variation in repeat number only occurs at loci 13 and 21, and therefore decreases the STRD that is observed. A Similar reduction in typing discriminatory power within defined bacterial lineages has been demonstrated with *M. tuberculosis* and the globally distributed Beijing clone where 12 or 15 loci typing schemes cannot differentiate between clusters within the lineage (Kremer *et al.*, 2005, Kato-Maeda *et al.*, 2011). Increasing the number of VNTR loci used in *M. tuberculosis* typing to 24 has however increased the discriminatory power of VNTR within the Beijing clone, where 24 loci could identify twice as many clusters of the *M. tuberculosis* Beijing clone compared to the 12 locus set (Jiao *et al.*, 2008).

To further investigate the discriminatory power of VNTR typing within the EMRSA-15 lineage, changes in LV and STRD between paired isolates were compared to more discriminatory SNP distances. Similar ranges of between 0 and 196 SNPs were demonstrated between zero, one and two locus variants between isolates. This would indicate a restricted range of SNPs differentiate EMRSA-15 isolates and further clarifies the high degree of clonality within the EMRSA-15 lineage. Some extent of this variation in SNPs between isolates may be accounted for by the cloud of diversity that has been demonstrated in *S. aureus* whereby individuals may be carrying heterogeneous populations of the bacterium. Different colonies from a single MRSA screening swab have been reported to differ from each other by up to 40 SNPs (Harris *et al.*, 2013, Paterson *et al.*, 2015), meaning each isolate in our study could be plus or minus 40 SNPs from the SNP difference values calculated.

Walker and colleagues have shown a similar result when comparing VNTR and SNPs with *M. tuberculosis* using a 24 locus scheme, where LVs of zero to two showed a range of zero to 200 SNPs between paired isolates. The authors also demonstrated that with increasing LV the range of SNPs remained constant but had higher upper and lower boundaries (LV of 10 ranging from 200 to 2000 SNPs for example) (Walker *et al.*, 2013). Somewhat conversely, Eyre *et al* also demonstrated increasing SNP difference with increasing LV using a seven locus system on over 300 *Clostridium difficile* isolates. The authors also showed that LV of zero and one between paired isolates only varied by up to 5 SNPs; a greatly reduced range compared to *S. aureus* or *M. tuberculosis*. This may be due to the biology of the organism, however *C. difficile* evolves faster than *M. tuberculosis* (1.5 SNPs per year (Didelot *et al.*, 2012) versus 0.5 SNPs per year (Walker *et al.*, 2013) respectively) but slower than *S aureus*, suggesting the difference may be in the discriminatory power of the *C. difficile* VNTR loci rather than a reduced mutation rate.

Both the Walker and Eyre studies have included isolates from different lineages in their respective bacteria and have demonstrated greater range of LV and STRD within the species. Here only a single highly clonal MRSA lineage was examined and accounts for the reduced variability in my VNTR data. VNTR discriminatory power may differ between MRSA lineages, as is seen with *M. tuberculosis* (Kato-Maeda *et al.*, 2011). MRSA ST239 is more a more diverse lineage compared to EMRSA-15 (Gray *et al.*, 2011, Hsu *et al.*, 2015) and therefore ST239 may have greater LV, STRD and SNP pairwise differences within the lineage.

VNTR typing data was also applied to my WGS phylogeny to determine how it reflected the Birmingham EMRSA-15 population structure. The mapping of the only

variable VNTR loci L13 and L21 to the WGS phylogeny showed a highly concordant EMRSA-15 population structure. Locus L13 was able to discriminating BHM and non BHM isolates with 100% concordance which further highlights the clonality of the BHM clade. Ahlstrom and colleagues investigated the epidemiology of 134 isolates of the veterinary pathogen *Mycobacterium avium* using a similar method (Ahlstrom *et al.*, 2015). The authors applied six polymorphic loci to their WGS phylogeny but found more limited clade-related mapping with VNTR repeat number. VNTR loci have been shown to have different allelic diversity and therefore different discriminatory power (Supply *et al.*, 2000, Supply *et al.*, 2006) which could mean the loci being investigated by Ahlstrom and colleagues are not variable enough to identify clones within *Mycobacterium avium* population (Ahlstrom *et al.*, 2015). Interestingly locus L21 is more variable than L13, due to smaller sized tandem repeats (Hardy *et al.*, 2006b) but was less concordant than L13 to the WGS phylogeny. This may be due to the genomic region of L13 having SNPs that would have been used to generate the phylogenetic tree.

4.5 The impact of fluoroquinolone usage on the Birmingham EMRSA-15 phylogeny

4.5.1 Fluoroquinolone use at HEFT

Both the Birmingham EMRSA-15 phylogeny and the use of fluoroquinolones at HEFT showed similar patterns that changed over the 25 period they were compared for.

Ciprofloxacin appears to have been introduced relatively late (1991) to HEFT, four years after it was licensed, however after 1991 ciprofloxacin use showed a similar trend to national usage with a steady increase between 1991 and 1996 (Livermore *et al.*,

2002). Total fluoroquinolone use increased sharply after 1996 when ofloxaxin was introduced and this drug was largely replaced with levofloxacin in 2003 due to the poor antibacterial efficacy of ofloxacin compared to levofloxacin (Davis and Bryson, 1994, Une *et al.*, 1988, George and Morrissey, 1997). This period of increase has been reported in other UK hospitals and has been followed by post 2003 reduction in use of fluoroquinolones both at HEFT and nationally (Aldeyab *et al.*, 2008, Sarma *et al.*, 2015) (Ashiru-Oredope *et al.*, 2012) (Cooke *et al.*, 2015) to help combat *Clostridium difficile* infection in UK hospitals. Post 2012 fluoroquinolone use at HEFT is also concordant with national data from the ESPAUR report (Public Health England, 2015) which states reduction in fluoroquinolone use has plateaued but levofloxacin usage is rising slowly. Therefore my data appears to be largely concordant with hospital prescribing patterns in the rest of the UK and is not demonstrating region or hospital specific biasing (an acute hospital trust for example). However several limitations should also be considered. Firstly ambiguity can arise around antibiotic usage data due to issues with how the data is collected and analysed. Antibiotics procured by wards may not all be administered therefore giving higher values for antibiotic usage than may actually be occurring but is likely to have a limited impact on annual data. National and local DDD values for drugs may also differ and can change over time, giving higher or lower DDDs than may have actually been used. This may partially account for the rapid spike in levofloxacin usage at HEFT in 2003 (Figure 19).

OBD data pre dating 2001 could be sourced, meaning antibiotic prescription data was determined as DDD and not DDD/OBD. The stable post 2001 OBD rate indicates very limited fluctuation in the in patient population at HEFT, suggesting the OBD rate prior to 2001 would have been similar to that seen between 2001 and 2006 (before the

merger with Good Hope Hospital). Antibiotic prescription data could only be sourced from a single hospital trust in Birmingham and QEHB and CHB may have had different prescribing policies and antibiotic usage over the 25 year time period. Antibiotic prescribing policy and stewardship is specific to each hospital trust in England and often reflects their specialism or function (Public Health England, 2015). For example QEHB is a trauma specialist trust and is likely to use more empirical prescribing of antibiotics compared to HEFT or CHB.

4.5.2 Antibiotic usage and phylogenetic comparison

The impact of antibiotic usage on the evolution of bacteria is now well established and the use of fluoroquinolone, third generation cephalosporin and macrolide antibiotics have been associated with the evolution of resistant MRSA lineages and increased risk of MRSA colonization (Charbonneau *et al.*, 2006) (Parienti *et al.*, 2011) (Tacconelli *et al.*, 2008). Resistance to fluoroquinolones has been proposed to facilitate the expansion of MRSA clones including EMRSA-15 but to date, no study has combined antibiotic usage over time with a time linked phylogenetic tree from a localized area or healthcare system.

Ciprofloxacin usage at HEFT started 1991 however both the phylogeny presented here and the Holden and colleagues data (Holden *et al.*, 2013) have shown that the emergence of EMRSA-15 and it's immediate evolution, the major branching of the tree between 1985 and 1990 (Figure 10), occurred prior to the introduction of ciprofloxacin into the HEFT trust in 1991. As discussed in section 4.3.2 the use of ciprofloxacin in clinical trials in the West Midlands may have resulted in this primary evolution and

major branching of the MCC tree between 1984 and 1990, but the ciprofloxacin usage at this time was in a minimal number of patients.

The gradual increase of ciprofloxacin between 1991 to 1997 followed by the more rapid increase in total fluoroquinolone use when ofloxacin was introduced in 1997 (causing a 4.5 fold increase in total fluoroquinolone use by 2002) coincided with the continual branching and expansion of the EMRSA-15 population in the Birmingham healthcare system. During this ten year time period the Birmingham EMRSA-15 population diversified more than any other time period in my analysis and may have been facilitated by increasing use of fluoroquinolones. From 1991 to 1997 the national EMRSA-15 population, bacteraemia rate and the national rate of fluoroquinolone use were also increasing (Livermore *et al.*, 2002) (Johnson *et al.*, 2001) (Ellington *et al.*, 2010). Therefore the expansion of EMRSA-15 in Birmingham coincides with the expansion of EMRSA-15 across the UK, as would be expected.

The evolution of the BHM clade (c1995) also occurred within this initial ten year period of ciprofloxacin use. Bayesian confidence intervals of the BHM clade emergence date overlap with the introduction of ofloxacin in 1997, indicating the possible selection pressure of increased ciprofloxacin and ofloxacin use may have precipitated the evolution of the BHM clade. Furthermore the BHM clade genomes contain a Pro453Ser mutation in the DNA supercoiling gene *grlB*, a mutation that is not frequently reported but has been showed to cause reduced MICs to ofloxacin (Takahashi *et al.*, 1998) in bacteria with *gyrA* Ser84Leu and *grlA* Ser80Phe mutations. While no literature performing a similar analysis could be sourced this theory of mutation, selection pressure and expansion echoes that of the primary EMRSA-15 evolution described by Holden and colleagues (Holden *et al.*, 2013). Similar to the increasing use of

ciprofloxacin and the evolution of ERMSA-15 in the late 1980s and early 1990s, ofloxacin use doubled annually between 1997 and 2003 allowing the BHM clade to expand and become established (comprising over 50% of Birmingham sampled isolates in 2003) before ofloxacin use was curtailed.

A further possible explanation for the dominance of the BHM clade is increased fitness from the alteration of DNA supercoiling gene structure. Mutations in DNA supercoiling genes have been demonstrated to cause a global fitness increase in Gram negative *Salmonella* Typhi and *Campylobacter jejuni* (Baker *et al.*, 2013, Han *et al.*, 2012). Therefore the *grlB* mutation in conjunction with the increasing fluoroquinolone usage when the clade emerged could have provided the BHM clade with a selective advantage over other EMRSA-15 strains, however this would require significant further study to be confirmed (see future work section).

Although my analysis has demonstrated a possible role of fluoroquinolone use in the evolution of EMRSA-15 the evolutionary impacts of IPC are difficult to quantify or assess (Lawes *et al.*, 2015). Screening of patient admissions (elective and emergency) are believed to have drastically reduced the MRSA burden in hospitals (Elgohari *et al.*, 2016). Reduced branching of the MCC tree was demonstrated from 2003 to 2010 when MRSA bacteriaemia rate was reducing, coinciding with a reduction in fluoroquinolone and a wide spectrum of IPC initiatives. Therefore it is likely a combination of enhanced IPC, screening and reduced fluoroquinolone use inducing the reduction in MRSA diversity.

Contrary to this pattern of reduced diversity with reducing fluoroquinolone usage, there was an increase in branching of the MCC tree from 2010 to 2015 when

fluoroquinolone use had plateaued. This is likely reflecting the increased number of outbreak and PII sampled isolates in my collection from this time, demonstrating small scale evolution occurring over restricted periods of time.

1.1 Desiccation tolerance of EMRSA-15

S*. aureus* is known to persist in the environment and one of the key factors contributing to environmental survival is tolerance to desiccation. Previous studies have examined desiccation tolerance between different MRSA sequence types or strains (Knight *et al.*, 2012, Baldan *et al.*, 2015, Duckworth and Jordens, 1990) however to date no previous study has compared EMRSA-15 sub-strains to each other or those from different geographical and climatological environments.

4.5.3 Desiccation assay

1.1.1.1 Methodology and control of desiccation chamber

The environmental conditions of a desiccation assay play an important role in generating results however few studies have actively controlled the environment the bacteria are being desiccated in. Here I adopted a similar methodology to that of Jawad and colleagues (Jawad *et al.*, 1996) and used a closed system with a controlled atmospheric RH in order to control environmental variables as much as possible to give robust results. Previous studies have shown that bacterial death rate is linked to RH and that RH is linked to temperature (Jawad *et al.*, 1996, Lidwell and Lowbury, 1950), indicating assays where RH and temperature are not controlled may produce variably

inaccurate results. Furthermore changing RH or temperature may in turn be affected by seasonal climate changes or differences between geographical locations of laboratories or even the availability of air conditioning (Rountree, 1963, Pettit and Lowbury, 1968). Validation of the desiccation chamber in my study showed there was a constant temperature but a variable relative humidity in an air-conditioned and temperature controlled laboratory; demonstrating the need for a controlled RH in the desiccation assay. Another variable that is likely to fluctuate is light intensity, which was not controlled in this experiment, however previous studies have shown no effect on desiccation in light or dark conditions (Beard-Pegler *et al.*, 1988). In reality the environment in the hospital would not be fixed, with daily fluctuations in light, temperature and humidity as was demonstrated when humidity and temperature readings were taken from a ward at HEFT. Furthermore assessing desiccation is difficult to evaluate without being affected by other stress responses such as starvation which are likely to induce the general SOS response, making a perfect model system difficult to achieve.

Only two studies have investigated the desiccation tolerance of EMRSA-15, both of which shared similar methodology. Studies by Baldan *et al* (Baldan *et al.*, 2015) and Knight *et al* (Knight *et al.*, 2012) used 200µl innocula of bacteria suspended in BHI broth plated onto empty sterile petri dishes and left in an open environment to desiccate. This may give higher bacterial recovery values due to the nutritious and protective environment provided by the BHI broth and may also be affected by dust or particulate matter landing on the petri dish from the open environment. By washing bacterial cells before desiccation and sealing them inside the desiccation chamber my results should

have controlled for these variables. These studies also set experiment end points at 120 hours (5 days) (Knight *et al.*, 2012) 168 hours (7 days) (Baldan *et al.*, 2015) whereas this study has continued to 35 days when bacteria were no longer recoverable; providing a greater understanding of the survival of EMRSA-15 over longer periods of time.

4.5.4 Isolate susceptibility to desiccation

Bacterial counts for each isolate varied over time, with each isolate showing dips and spikes in enumerated bacteria at each time point and no isolate showing a directly linear reduction in count over time. Previous studies have shown similar variability in bacterial counts over time both in *S. aureus* (Knight *et al.*, 2012, Baldan *et al.*, 2015, Jawad *et al.*, 1996) and other species (Jawad *et al.*, 1996) and is likely due to variation in the number of bacteria inoculated onto coverslips and the inherent inaccuracy of counting bacteria. Clumping of *S. aureus* during washing could mean an uneven number of bacteria were inoculated onto each coverslip (variation in number of bacteria on each coverslip was present) or clumping when eluting from coverslips or application to plates via capillaries in the spiral plater could also give increased counts on plates. Clumped bacteria may survive desiccation better due to a reduced net loss of water, giving higher survival rates at a time point after a previously lower count from an enumerated coverslip with no clumped bacteria. Furthermore human error in pipetting, spiral plating or counting of bacterial colonies can also add variation to results.

Of the ten EMRSA-15 isolates analysed the Hong Kong isolate HK_N5 showed the highest tolerance to desiccation and CT300 the lowest tolerance when suspended in water. From our phylogenetic analysis the Hong Kong isolate is the most genetically distant isolate compared to the other EMRSA-15 isolates in the assay and has possible undergone selective evolution in the Singapore and Hong Kong healthcare environments it has passed through (section 3.2.3). It is possible that the strain has evolved to have a higher tolerance of desiccation but no genetic evidence could be found for this in the WGS analysis however epigenetic changes to the transcriptome could account for the increased tolerance and were not investigated here.

When suspended in water the BHM isolates showed slightly higher desiccation tolerance (higher range of AUC values and reduced slope gradients) compared to non-BHM isolates however this was ablated when suspended in BSA where AUC values and linear regression gradients overlapped. This would suggest that desiccation tolerance is not heightened in the BHM clade and is unlikely to be a reason for the strains success in Birmingham. Increasing the number of BHM and non BHM isolates may give further distinction to these results.

The poor desiccation tolerance of isolates CT300 is likely due to its impaired ability to produce staphyloxanthin. This was evident from the lack of pigmentation of the bacterial colonies and WGS data that showed a SNP in the *ctrN* gene comprising part of the *ctrNOPQM* operon that codes for the production of staphyloxanthin (Pelz *et al.*, 2005). Staphyloxanthin is known to be an important cellular mechanism for desiccation tolerance due to it protecting from oxidative stress along with neutrophil killing in host

infection (Liu and Nizet, 2009, Pelz *et al.*, 2005). Two successful MRSA lineages also lack the production of staphyloxanthin. Firstly an early branching CA-MRSA lineage in clonal complex 75 that lacks staphyloxanthin has been reported to be the dominant MRSA lineage among indigenous Australian communities (McDonald *et al.*, 2006, Holt *et al.*, 2011). This strain has a predilection for skin and soft tissue infection and is rarely identified in hospitals (Brennan *et al.*, 2013) which Tong *et al* 2013 suggest is due to the strain being less fit for the hospital environment compared to staphyloxanthin producing strains. On the other hand EMRSA-16 is a successful HA-MRSA lineage found mainly in Southern England that emerged at a similar time to EMRSA-15, and is also ciprofloxacin resistant, which has a nonsense mutation in *ctrM* to give non-pigmented colonies (McAdam *et al.*, 2012). EMRSA-16 is not found in the community and McAdam *et al* 2012 (McAdam *et al.*, 2012) account this to the lack of staphyloxanthin (along with other virulence factors such as *hla* and *agr*) giving the strain reduced virulence and an inability to infect healthy individuals outside of the hospital. Both of these examples demonstrate staphyloxanthin lacking strains have reduced fitness compared to staphyloxanthin producing strains (such as EMRSA-15) but suggest the pigment is not essential for maintenance in either the community or the hospital setting which in turn may indicate desiccation tolerance has a minor role in strain fitness.

4.5.5 Effect of protein suspension media

To date no study has examined the desiccation tolerance of EMRSA-15 when suspended in BSA. A statistically significant increase in the desiccation tolerance of all *S. aureus*

strains was shown when bacteria were suspended in BSA. A solution of 3% w/v was chosen due to this being the percentage of protein present in "dirty environmental conditions" in an attempt to emulate an unclean hospital environment (Tuladhar *et al.*, 2012, Kawamura-Sato *et al.*, 2008). These finding has been echoed in previous literature in both *S. aureus* (Chaibenjawong and Foster, 2010) and *A. baumanii* (Jawad *et al.*, 1996) where a significant increase in long term survival and a reduced death rate of both organisms was observed. The exact reasoning behind the increased desiccation tolerance remains poorly understood, however the protein may form a coating around the bacteria, preventing loss of water from the cell and reducing UV damage (Fraud *et al.*, 2001).

The increase in desiccation tolerance over the first 24 hours of the experiment and increased duration of survival has implications in the hospital environment and suggests the bacteria can survive longer in the dirty hospital environment. Contamination of hospital environments is known to cause an increased risk of infection with MRSA (and other HA pathogens) (Chemaly *et al.*, 2014, Otter *et al.*, 2011) and transmission from environment to patient is a known route of MRSA infection (Tajeddin *et al.*, 2016, Boyce, 2007). Hospital wards contaminated with proteins from blood, faeces or skin particles may therefore act an extended reservoir of viable bacteria to be transmitted. Furthermore BSA has been shown to provide protection to biocide agents (Fraud *et al.*, 2001, Simões *et al.*, 2006) and a combination of desiccation and biocide tolerance provided by organic proteins in the environment may allow MRSA to withstand standard hospital cleaning practices. Recent changes in hospital cleaning practices due to the inefficiency of cleaning regimes mean more modern "no-touch" cleaning technologies such as UV irradiation and gaseous hydrogen peroxide are being

employed to decontaminate wards and side rooms. Studies combining desiccation tolerance with UV irradiation or hydrogen peroxide exposure may give further insight into MRSA desiccation tolerance in the moden hospital environment (Carling and Bartley, 2010, Boyce, 2016).

Interestingly all of the EMRSA-15 isolates showed poorer desiccation tolerance than the MSSA control strain in BSA but not when suspended in water. Previous studies have investigated the desiccation tolerance of *S. aureus* with the gain or loss of resistance to antibiotics other than fluoroquinlones (Chaibenjawong and Foster, 2010, Beard-Pegler *et al.*, 1988, Duckworth and Jordens, 1990). Lacey *et al* 1972 (Lacey, 1972) demonstrated no effect on desiccation tolerance with the loss or gain of antibiotic resistance to methicillin, penicillin or tetracycline, however their desiccation exposure time was only 6 hours (an end time point before the first sampling at 12 hours in our study), and similarly Duckworth and Jordens et al 1990 (Duckworth and Jordens, 1990) demonstrated that isolates with increased number of resistances tolerated desiccation similar to those with fewer resistances. None of these studies included EMRSA-15 isolates and conversely studies investigating desiccation tolerance of EMRSA-15 have not included MSSA isolates. The data presented here suggests EMRSA-15 is poorer at surviving desiccation that MSSA in a dirty environment, however the lack of other MRSA STs in the study and only a single MSSA control isolate (no screening or clinical MSSA isolates) means these result may be biased.

4.5.6 Effect of ciprofloxacin resistance

Antibiotic resistance is thought to play an important role in the overall fitness of HA-MRSA in the hospital environment due to the high usage of a myriad of antibiotics in this ecological niche.

By including the historical CHB isolate (CT124) from the progenitor population of EMRSA-15 I have been able to compare the effect of the *gyrA* and *grlA* mutations on the EMRSA-15 lineages tolerance to desiccation. Overall, the CT124 isolates had comparable desiccation tolerance compared to the ciprofloxacin resistant isolates, having the second highest AUC values in water but the lowest number of bacteria surviving 576 hours (with the exception of CT300) as well as median range values when suspended in BSA but was again the poorest surviving strain post 408 hours. Mutation in *gyrA* in *Salmonella enterica* has been shown to cause a change in transcription leading to increased expression of stress response pathways (Webber *et al.*, 2013), indicating DNA supercoiling gene alteration (and therefore fluoroquinolone resistance) can affect the overall fitness of the bacteria, however there has been no similar study in Gram positive organisms. My study suggests fluoroquinolone resistance appears to have given a small increase in desiccation tolerance, which is controlled by stress response, indicating possible fitness advantage from the change in DNA supercoiling gene structure that induces fluoroquinolone resistance. The poorer desiccation tolerance of the staphyloxanthin lacking CT300 isolate would however suggest cellular environmental tolerance mechanisms provide a greater degree of protection than potential increased stress response activity provided by fluoroquinolone resistance. Repetition of the desiccation assay with targeted point mutation in *grlA* and *gyrA* mutants of isolate CT124 may demonstrate this increased tolerance more appropriately along with additional ciprofloxacin sensitive ST22 isolates.

1.2 Conclusions

By constructing a bespoke bioinformatics pipeline and applying Bayesian phylogenetics to 437 globally sampled EMRSA-15 genomes I have expanded the current knowledge base of EMRSA-15 evolution and epidemiology by determining several key findings.

The identification, sequencing and phylogenetic analysis of a single isolate sampled from c1985 has shown that a hospital associated ciprofloxacin sensitive ST22 population (lacking the Ser80Phe mutation in *grlA* and Ser84Leu in *gyrA* mutation) that was highly related to current EMRSA-15 was circulating in the West Midlands during the proposed period fluoroquinolone resistant ST22 (EMRSA-15) emerged. This would indicate an EMRSA-15 progenitor population was present in the West Midlands during the mid 1980s; the same time that my Bayesian analysis indicates EMRSA-15 first emerged (c1984). Therefore I have been able to support the hypothesis suggested by Holden and colleagues that EMRSA-15 emerged in the West Midlands but have arrived at an emergence date two years prior to the respective authors proposal of 1986.

Geographically restricted clones of EMRSA-15 in Birmingham, Guernsey and Hong Kong were demonstrated from the global phylogeny.

Using a time linked MCC phylogenetic tree I have identified a clonal, highly localized and dominant EMRSA-15 clone (BHM clone) that emerged in the mid 1990s which was present in the three largest hospitals in Birmingham. Further WGS analysis has shown the BHM clone to have a unique SNP in the *grlB* (topoisomerase IV subunit B)

gene which may give the clone a selective advantage in the Birmingham healthcare system.

Guernsey and Hong Kong both represent island populations of EMRSA-15 and demonstrated localized evolution in their respective geographical regions. The population in Guernsey demonstrated a highly localized strain was present alongside EMRSA-15 strains related to the UK mainland. This is likely due to an importation of EMRSA-15 from the UK followed by clonal expansion, alongside continuous sporadic importations from UK health workers and patients trafficking between Guernsey and the UK.

In Hong Kong ST239 dominates and EMRSA-15 is a minority clone. Here I have shown that a group of related EMRSA-15 isolates across four Hong Kong hospitals have a single ancestral origin from Singapore, demonstrating transmission of EMRSA-15 within Hong Kong but also between Asian countries. This further highlights that global population movements can promote the spread of pathogens. The combination of Birmingham, Guernsey and Hong Kong epidemiology presented here along with evidence of UK localization of EMRSA-15 sub strains demonstrated by Reuter and colleagues is indicative of patient and healthcare worker movement being highly influential in local, national and international EMRSA-15 population structuring.

All versus all pairwise comparison of the 210 EMRSA-15 isolates that were VNTR typed and whole genome sequenced showed a limited range in LV, STRD and SNPs between isolates in this collection. This reinforces the highly clonal nature of EMRSA-15. Mapping of two polymorphic VNTR loci to the MCC phylogeny showed, VNTR and WGS (Bayesian

analysis) to be concordant for the Birmingham EMRSA-15 population with VNTR typing discriminating the BHM clade with 100% concordance at a single locus.

By performing a unique comparative analysis of the Birmingham MCC phylogeny with antibiotic prescription data from HEFT I have demonstrated a relationship between increasing fluoroquinolone use and evolution of EMRSA-15. Increasing ciprofloxacin use and the introduction of ofloxacin in the 1990s coincided with the emergence and expansion of the BHM clade containing a Pro456Ser *grlB* mutation. This increase in fluoroquinolone use may have provided a selection pressure for the evolution and expansion of the EMRSA-15 sub lineage. This demonstrates the introduction of new drug classes may influence localized EMRSA-15 population structures and therefore change the dynamic of hospital MRSA burden.

Finally using carefully controlled methodology I have demonstrated a significant increase in EMRSA-15 desiccation tolerance when isolates were suspended in protein media emulating dirty hospital environmental conditions. This would suggest that EMRSA-15 can survive longer in dirty hospital environments, therefore allowing surfaces to act as reservoirs for infection over increased periods of time. Examination of multiple EMRSA-15 sub-lineage isolates has shown limited difference in desiccation tolerance between BHM and non-BHM isolates when suspended in water but not in BSA. This would suggest changes in osmotic stress tolerance may give increased desiccation tolerance in BHM isolates (due to the reduction in osmotic stress tolerance when suspended in BSA).

1.3 Future Work

- To further analyse the impact of the *grlB* mutation on the fitness of the BHM clade by:
    - Generating Pro453Ser *grlB* point mutations in ciprofloxacin sensitive progenitor isolates and determine impacts on MIC to fluoroquinolones.
    - Using fitness experiments such as co-culture growth assays, pH tolerance or fatty acid resistance to identify any further physiological characteristics that may provide the BHM clade with a selective advantage.

- To investigate any possible differences in biocide tolerance, triclosan and chlorhexidine between sub-strains of EMRSA-15 including the presence or absence of *qac* genes and changes in MIC or minimum bactericidal concentrations (MBC).

- To use the Oxford Nanopore minION sequencer to sequence the MGE and VNTR long repeat regions, that were excluded from the analysis presented here, to link mini satellite regions to possible phenotypic characteristic changes and determine the repertoire of MGEs circulating in the Birmingham EMRSA-15 population.

5 Appendices

Table 25: Isolates and associated metadata that were whole genome sequenced and included in the study. Entries colour red have been identified as the BHM clone.

Isolate Identifier	Country	Hospital Trust	Collection	Screening or clinical sample	Sampling Date	VNTR profile	L1	L13	L15	L16	L21	L5	L7	Sequence type	Ciprofloxacin resistance	Coverage
A11	UK	HEFT	HEFT ITU	Screening	04/02/2002	1 3 0 3 17 - 2	1	3	0	3	17	-	2	22	R	15.43
A33	UK	HEFT	HEFT ITU	Screening	11/02/2002	1 4 0 3 16 - 2	1	4	0	3	16	-	2	22	R	15.76
A120	UK	HEFT	HEFT ITU	Screening	27/02/2002	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	22.45
A174	UK	HEFT	HEFT ITU	Screening	13/03/2002	1 4 0 3 13 - 2	1	4	0	3	13	-	2	22	R	15.08
A193	UK	HEFT	HEFT ITU	Screening	15/03/2002	1 4 0 3 17 - 2	1	4	0	3	17	-	2	22	R	15.71
A221	UK	HEFT	HEFT ITU	Screening	25/03/2002	1 4 0 3 13 - 2	1	4	0	3	13	-	2	22	R	15.79
A280	UK	HEFT	HEFT ITU	Screening	29/04/2002	1 4 0 3 13 - 2	1	4	0	3	13	-	2	22	R	19.60
A328	UK	HEFT	HEFT ITU	Screening	24/05/2002	1 3 0 3 17 - 2	1	3	0	3	17	-	2	22	R	17.97
A442	UK	HEFT	HEFT ITU	Screening	14/08/2002	1 4 0 3 17 - 2	1	4	0	3	17	-	2	22	R	16.12
A453	UK	HEFT	HEFT ITU	Screening	30/08/2002	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	17.84
A510	UK	HEFT	HEFT ITU	Screening	27/09/2002	1 4 0 3 17 - 2	1	4	0	3	17	-	2	22	R	15.69
A540	UK	HEFT	HEFT ITU	Screening	30/09/2002	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	20.73
A556	UK	HEFT	HEFT ITU	Screening	02/07/2003	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	16.51
A640	UK	HEFT	HEFT ITU	Screening	06/08/2003	1 2 0 3 16 - 2	1	2	0	3	16	-	2	22	R	19.40
A714	UK	HEFT	HEFT ITU	Screening	20/08/2003	1 4 0 3 16 - 2	1	4	0	3	16	-	2	22	R	24.30
A746	UK	HEFT	HEFT ITU	Screening	20/08/2003	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	17.89

A780	UK	HEFT	HEFT ITU	Screening	01/09/2003	1 4 0 3 16 - 2	1	4	0	3	17	-	2	22	R	16.97
A782	UK	HEFT	HEFT ITU	Screening	01/09/2003	1 3 0 3 16 - 2	1	3	0	3	16	-	2	22	R	15.34
A811	UK	HEFT	HEFT ITU	Screening	05/09/2003	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	18.83
A817	UK	HEFT	HEFT ITU	Screening	12/09/2003	1 2 0 3 16 - 2	1	2	0	3	16	-	2	22	R	16.44
A878	UK	HEFT	HEFT ITU	Screening	01/10/2003	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	15.96
A901	UK	HEFT	HEFT ITU	Screening	08/10/2003	1 3 0 3 16 - 2	1	2	0	3	18	-	2	22	R	15.97
A958	UK	HEFT	HEFT ITU	Screening	31/10/2003	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	15.58
A1052	UK	HEFT	HEFT ITU	Screening	12/11/2003	1 2 0 3 16 - 2	1	2	0	3	16	-	2	22	R	17.60
A1043	UK	HEFT	HEFT ITU	Screening	14/11/2003	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	15.16
A1039	UK	HEFT	HEFT ITU	Screening	17/11/2003	1 5 0 3 17 - 2	1	5	0	3	17	-	2	22	R	19.41
A1093	UK	HEFT	HEFT ITU	Screening	05/12/2003	1 4 0 3 17 - 2	1	4	0	3	17	-	2	22	R	15.87
A1099	UK	HEFT	HEFT ITU	Screening	05/12/2003	1 3 0 3 17 - 2	1	3	0	3	17	-	2	22	R	15.33
A1103	UK	HEFT	HEFT ITU	Screening	05/12/2003	1 3 0 3 17 - 2	1	3	0	3	17	-	2	22	R	15.64
A1173	UK	HEFT	HEFT ITU	Screening	02/01/2004	1 2 0 3 15 - 2	1	2	0	3	15	-	2	22	R	15.98
A1247	UK	HEFT	HEFT ITU	Screening	23/01/2004	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	15.90
A1257	UK	HEFT	HEFT ITU	Screening	02/02/2004	1 4 0 3 17 - 2	1	4	0	3	17	-	2	22	R	26.12
A1263	UK	HEFT	HEFT ITU	Screening	04/02/2004	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	17.50
A1298	UK	HEFT	HEFT ITU	Screening	18/02/2004	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	15.14
A1316	UK	HEFT	HEFT ITU	Screening	20/02/2004	1 4 0 3 17 - 2	1	4	0	3	17	-	2	22	R	20.12
A1323	UK	HEFT	HEFT ITU	Screening	23/02/2004	1 4 0 3 17 - 2	1	4	0	3	17	-	2	22	R	15.56
R198	UK	PEH	Guernsey	Unknown	06/06/2005	1 4 0 3 17 - 2	1	4	0	3	17	-	2	22	R	44.28
EMM4	UK	HEFT	HEFT surgical	Screening	13/11/2005	1 2 0 3 16 - 2	1	2	0	3	16	-	2	22	R	25.11
EMT8	UK	HEFT	HEFT surgical	Screening	19/11/2005	1 4 0 3 14 - 2	1	4	0	3	14	-	2	22	R	194.00
EMT10	UK	HEFT	HEFT surgical	Screening	22/11/2005	1 4 0 3 16 - 2	1	4	0	3	16	-	2	22	R	83.78
EMM23	UK	HEFT	HEFT surgical	Screening	01/12/2005	1 2 0 3 10 - 2	1	2	0	3	10	-	2	22	R	26.75
EMM38	UK	HEFT	HEFT surgical	Screening	12/12/2005	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	15.71
EMM63	UK	HEFT	HEFT surgical	Screening	22/12/2005	1 4 0 3 16 - 2	1	4	0	3	16	-	2	22	R	27.50

EMM88	UK	HEFT	HEFT surgical	Screening	14/01/2006	1 4 0 3 17 - 2	1	4	0	3	17	-	2	22	R	97.59
EMM138	UK	HEFT	HEFT surgical	Screening	30/01/2006	1 2 0 3 18 - 2	1	2	0	3	18	-	2	22	R	111.24
EMM139	UK	HEFT	HEFT surgical	Screening	06/02/2006	1 2 0 3 19 - 2	1	2	0	3	18	-	2	22	R	70.89
EMM144	UK	HEFT	HEFT surgical	Screening	08/02/2006	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	108.22
EMM146	UK	HEFT	HEFT contemporary	Screening	08/02/2006	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	116.00
EMM265	UK	HEFT	HEFT surgical	Screening	27/03/2006	1 2 0 3 16 - 2	1	2	0	3	16	-	2	22	R	71.01
EMT154	UK	HEFT	HEFT surgical	Screening	09/04/2006	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	57.06
EMM302	UK	HEFT	HEFT surgical	Screening	12/04/2006	1 2 0 2 13 - 2	1	2	0	2	13	-	2	22	R	16.37
EMM320	UK	HEFT	HEFT surgical	Screening	14/04/2006	1 2 0 3 12 - 2	1	2	0	3	12	-	2	22	R	182.00
EMT194	UK	HEFT	HEFT surgical	Screening	01/05/2006	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	17.53
EMM419	UK	HEFT	HEFT contemporary	Screening	22/05/2006	1 5 0 3 12 - 2	1	5	0	3	12	-	2	22	R	53.54
EMM418	UK	HEFT	HEFT surgical	Screening	25/05/2006	1 2 0 3 16 - 2	1	2	0	3	16	-	2	22	R	57.13
EMM443	UK	HEFT	HEFT surgical	Screening	02/06/2006	1 4 0 3 17 - 2	1	4	0	3	17	-	2	22	R	421.78
EMM451	UK	HEFT	HEFT surgical	Screening	06/06/2006	1 3 0 3 17 - 2	1	3	0	3	17	-	2	22	R	170.62
EMM450	UK	HEFT	HEFT surgical	Screening	10/06/2006	1 2 0 3 14 - 2	1	2	0	3	14	-	2	22	R	107.23
EMM513	UK	HEFT	HEFT surgical	Screening	25/06/2006	1 1 0 3 17 - 2	1	1	0	3	17	-	2	22	R	45.95
EMM573	UK	HEFT	HEFT surgical	Screening	15/07/2006	1 4 0 3 17 - 2	1	4	0	3	17	-	2	22	R	129.41
EMT309	UK	HEFT	HEFT surgical	Screening	22/08/2006	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	51.10
EMT311	UK	HEFT	HEFT surgical	Screening	22/08/2006	1 2 0 3 16 - 2	1	2	0	3	16	-	2	22	R	27.76
ECMM082	UK	HEFT	HEFT surgical	Screening	04/10/2006	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	170.64
ECMM102	UK	HEFT	HEFT surgical	Screening	19/10/2006	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	23.79
ECMM116	UK	HEFT	HEFT surgical	Screening	04/11/2006	1 4 0 2 17 - 2	1	4	0	2	17	-	2	22	R	17.90
ECMM195	UK	HEFT	HEFT surgical	Screening	16/12/2006	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	192.87
ECMM265	UK	HEFT	HEFT surgical	Screening	12/02/2007	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	41.06
R012	UK	PEH	Guernsey	Screening	28/03/2007	1 3 0 3 16 - 2	1	3	0	3	16	-	2	22	R	26.06
ECMM349	UK	HEFT	HEFT surgical	Screening	27/04/2007	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	46.32
R245	UK	PEH	Guernsey	Unknown	17/07/2010	1 4 0 3 16 - 2	1	4	0	3	16	-	2	22	R	15.04
R874	UK	HEFT	HEFT contemporary	Screening	06/06/2012	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	105.08

CT273	UK	PEH	Guernsey	Clinical	13/09/2012	1 4 0 3 18 - 2	1	4	0	3	18	-	2	22	R	32.09
CT275	UK	PEH	Guernsey	Clinical	07/01/2013	1 4 0 3 18 - 2	1	4	0	3	18	-	2	22	R	26.10
R841	UK	HEFT	HEFT contemporary	Screening	20/03/2013	1 2 0 3 15 2 2	1	4	0	3	15	2	2	22	R	103.31
R853	UK	HEFT	HEFT contemporary	Screening	08/05/2013	1 2 0 3 21 - 2	1	2	0	3	21	-	2	22	R	103.75
R862	UK	HEFT	HEFT contemporary	Screening	20/05/2013	1 2 0 3 19 - 2	1	4	0	3	19	-	2	22	R	83.36
R863	UK	HEFT	HEFT contemporary	Screening	20/05/2013	1 2 0 3 15 - 2	1	2	0	3	15	-	2	22	R	93.98
R869	UK	HEFT	HEFT contemporary	Screening	21/05/2013	1 2 0 3 18 - 2	1	2	0	3	18	-	2	22	R	78.17
CT268	UK	PEH	Guernsey	Screening	24/06/2013	1 5 0 3 18 - 2	1	5	0	3	18	-	2	22	R	23.40
R885	UK	HEFT	HEFT contemporary	Screening	29/06/2013	1 4 0 4 14 - 2	1	4	0	4	14	-	2	22	R	112.97
R890	UK	HEFT	HEFT contemporary	Screening	25/07/2013	1 5 0 3 17 - 2	1	5	0	3	17	-	2	22	R	111.20
R891	UK	HEFT	HEFT contemporary	Screening	25/07/2013	1 5 0 3 17 - 2	1	5	0	3	17	-	2	22	R	107.25
CT276	UK	PEH	Guernsey	Clinical	22/09/2013	1 4 0 3 17 - 2	1	4	0	3	17	-	2	22	R	23.11
CT139	UK	HEFT	HEFT contemporary	Clinical	10/10/2013	1 2 0 3 14 - 2	1	2	0	3	14	-	2	22	R	17.41
CT140	UK	HEFT	HEFT contemporary	Clinical	15/10/2013	1 2 0 3 14 - 2	1	2	0	3	14	-	2	22	R	16.28
CT145	UK	HEFT	HEFT contemporary	Screening	28/10/2013	1 4 0 3 17 - 2	1	4	0	3	17	-	2	22	R	16.63
CT146	UK	HEFT	HEFT contemporary	Screening	31/10/2013	1 2 0 2 18 - 2	1	2	0	2	18	-	2	22	R	16.43
CT143	UK	HEFT	HEFT contemporary	Clinical	03/11/2013	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	20.68
CT147	UK	HEFT	HEFT contemporary	Clinical	04/11/2013	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	22.90
CT148	UK	HEFT	HEFT contemporary	Clinical	07/11/2013	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	23.31
CT149	UK	HEFT	HEFT contemporary	Clinical	10/11/2013	1 4 0 3 17 - 2	1	4	0	3	17	-	2	22	R	15.79
CT204	UK	CHB	CHB Contemporary	Screening	18/11/2013	1 3 0 3 16 - 2	1	3	0	3	16	-	2	22	R	22.46
CT200	UK	QEHB	QEHB Contemporary	Screening	19/11/2013	1 5 0 3 18 - 2	1	5	0	3	18	-	2	22	R	22.18
CT209	UK	CHB	CHB Contemporary	Clinical	27/11/2013	1 4 0 3 17 - 2	1	4	0	3	17	-	2	22	R	15.17
CT211	UK	CHB	CHB Contemporary	Screening	01/12/2013	1 4 0 3 17 - 2	1	2	0	3	17	-	2	22	R	20.19
CT154	UK	HEFT	HEFT contemporary	Clinical	03/12/2013	1 3 0 3 20 - 2	1	3	0	3	20	-	2	22	R	27.18
CT210	UK	CHB	CHB Contemporary	Clinical	04/12/2013	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	20.53
CT207	UK	CHB	CHB Contemporary	Screening	05/12/2013	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	18.68
CT206	UK	CHB	CHB Contemporary	Clinical	06/12/2013	1 2 0 3 14 - 2	1	2	0	3	14	-	2	22	R	17.02

CT199	UK	QEHB	QEHB Contemporary	Screening	07/12/2013	1 3 0 3 17 - 2	1	3	0	3	17	-	2	22	R	25.81
CT155	UK	HEFT	HEFT contemporary	Screening	09/12/2013	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	15.00
CT213	UK	CHB	CHB Contemporary	Screening	10/12/2013	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	17.63
CT202	UK	QEHB	QEHB Contemporary	Screening	11/12/2013	1 3 0 3 14 - 2	1	3	0	3	14	-	2	22	R	15.75
CT158	UK	HEFT	HEFT contemporary	Screening	12/12/2013	1 4 0 3 17 - 2	1	2	0	3	17	-	2	22	R	18.57
CT156	UK	HEFT	HEFT contemporary	Screening	13/12/2013	1 4 0 3 17 - 2	1	4	0	3	17	-	2	22	R	15.62
CT165	UK	HEFT	HEFT contemporary	Screening	13/12/2013	1 4 0 3 17 - 2	1	4	0	3	17	-	2	22	R	15.81
CT166	UK	HEFT	HEFT contemporary	Screening	15/12/2013	1 4 0 3 15 - 2	1	4	0	3	15	-	2	22	R	16.88
CT212	UK	CHB	CHB Contemporary	Clinical	18/12/2013	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	50.40
CT168	UK	HEFT	HEFT contemporary	Clinical	19/12/2013	1 3 0 2 15 - 2	1	3	0	2	15	-	2	22	R	19.73
CT180	UK	QEHB	QEHB Contemporary	Screening	19/12/2013	1 2 0 3 18 - 2	1	2	0	3	18	-	2	22	R	52.24
CT169	UK	HEFT	HEFT contemporary	Screening	20/12/2013	1 5 0 3 17 - 2	1	5	0	3	17	-	2	22	R	17.70
CT170	UK	HEFT	HEFT contemporary	Screening	24/12/2013	1 2 0 3 15 - 2	1	2	0	3	15	-	2	22	R	18.52
CT135	UK	HEFT	HEFT contemporary	Screening	28/12/2013	1 2 0 3 12 - 2	1	2	0	3	12	-	2	22	R	15.71
CT136	UK	HEFT	HEFT contemporary	Screening	30/12/2013	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	18.06
CT159	UK	HEFT	HEFT contemporary	Screening	31/12/2013	1 2 0 3 20 - 2	1	2	0	3	18	-	2	22	R	18.57
CT171	UK	HEFT	HEFT contemporary	Screening	01/01/2014	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	15.30
CT141	UK	HEFT	HEFT contemporary	Screening	05/01/2014	1 3 0 3 17 - 2	1	3	0	3	17	-	2	22	R	15.13
CT160	UK	HEFT	HEFT contemporary	Screening	06/01/2014	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	23.36
CT161	UK	HEFT	HEFT contemporary	Screening	07/01/2014	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	16.03
CT163	UK	HEFT	HEFT contemporary	Screening	13/01/2014	1 2 0 2 15 - 2	1	2	0	2	15	-	2	22	R	15.72
CT164	UK	HEFT	HEFT contemporary	Screening	14/01/2014	1 3 0 3 15 - 2	1	3	0	3	15	-	2	22	R	21.33
CT181	UK	QEHB	QEHB Contemporary	Clinical	24/01/2014	1 3 0 3 17 - 2	1	3	0	3	17	-	2	22	R	46.34
CT190	UK	QEHB	QEHB Contemporary	Clinical	30/01/2014	1 4 0 3 16 - 2	1	4	0	3	16	-	2	22	R	19.93
CT182	UK	QEHB	QEHB Contemporary	Screening	01/02/2014	1 2 0 3 13 - 2	1	3	0	3	13	-	2	22	R	18.83
CT183	UK	QEHB	QEHB Contemporary	Screening	01/02/2014	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	15.24
CT175	UK	QEHB	QEHB Contemporary	Screening	04/02/2014	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	23.52

CT196	UK	QEHB	QEHB Contemporary	Screening	11/02/2014	1 4 0 3 18 - 2	1	4	0	3	18	-	2	22	R	15.49
CT218	UK	CHB	CHB Contemporary	Screening	12/02/2014	1 3 0 3 17 - 2	1	3	0	3	17	-	2	22	R	53.25
R979	UK	HEFT	HEFT contemporary	Screening	13/02/2014	1 2 0 3 17 2 2	1	2	0	3	17	2	2	22	R	107.11
CT219	UK	CHB	CHB Contemporary	Screening	13/02/2014	1 3 0 3 17 - 2	1	3	0	3	17	-	2	22	R	26.81
CT220	UK	CHB	CHB Contemporary	Screening	13/02/2014	1 3 0 3 17 - 2	1	3	0	3	17	-	2	22	R	18.85
CT203	UK	QEHB	QEHB Contemporary	Screening	15/02/2014	1 2 0 3 10 - 2	1	2	0	3	10	-	2	22	R	19.09
CT230	UK	CHB	CHB Contemporary	Screening	18/02/2014	1 4 0 3 16 - 2	1	4	0	3	16	-	2	22	R	20.87
CT227	UK	CHB	CHB Contemporary	Screening	20/02/2014	1 2 0 3 16 - 2	1	2	0	3	16	-	2	22	R	23.00
CT225	UK	CHB	CHB Contemporary	Screening	22/02/2014	1 4 0 3 10 - 2	1	4	0	3	10	-	2	22	R	21.34
CT229	UK	CHB	CHB Contemporary	Screening	25/02/2014	1 4 0 3 10 - 2	1	4	0	3	10	-	2	22	R	99.69
CT236	UK	CHB	CHB Contemporary	Screening	09/03/2014	1 4 0 3 16 - 2	1	4	0	3	16	-	2	22	R	19.81
CT244	UK	CHB	CHB Contemporary	Screening	15/03/2014	1 2 0 3 11 - 2	1	4	2	3	11	-	2	22	R	15.18
R1032	UK	HEFT	HEFT contemporary	Screening	10/04/2014	1 4 0 3 17 - 2	1	4	0	3	17	-	2	22	R	373.32
R988	UK	HEFT	HEFT contemporary	Clinical	15/04/2014	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	117.54
CT286	UK	QEHB	QEHB Contemporary	Clinical	04/05/2014	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	47.62
CT287	UK	QEHB	QEHB Contemporary	Screening	06/06/2014	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	40.12
CT279	UK	PEH	Guernsey	Clinical	24/06/2014	1 4 0 3 18 - 2	1	4	0	3	18	-	2	22	R	15.31
CT288	UK	QEHB	QEHB Contemporary	Clinical	25/06/2014	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	36.79
R1011	UK	HEFT	HEFT contemporary	Screening	08/07/2014	1 3 0 3 17 - 2	1	3	0	3	17	-	2	22	R	293.58
R1012	UK	HEFT	HEFT contemporary	Screening	10/07/2014	1 3 0 3 17 - 2	1	3	0	3	17	-	2	22	R	264.03
CT289	UK	QEHB	QEHB Contemporary	Clinical	11/07/2014	1 2 0 3 14 - 2	1	2	0	3	14	-	2	22	R	16.67
R1013	UK	HEFT	HEFT contemporary	Screening	18/07/2014	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	374.65
R1029	UK	HEFT	HEFT contemporary	Screening	24/07/2014	1 3 0 3 17 - 2	1	3	0	3	17	-	2	22	R	316.05
R1028	UK	HEFT	HEFT contemporary	Screening	28/07/2014	1 3 0 3 17 - 2	1	3	0	3	17	-	2	22	R	270.34
R1039	UK	HEFT	HEFT contemporary	Screening	23/08/2014	1 3 0 3 12 - 2	1	3	0	3	12	-	2	22	R	428.79
R1043	UK	HEFT	HEFT contemporary	Clinical	04/09/2014	1 3 0 3 17 - 2	1	3	0	3	17	-	2	22	R	336.41
CT291	UK	QEHB	QEHB Contemporary	Screening	22/09/2014	1 2 0 3 16 - 2	1	2	0	3	16	-	2	22	R	15.37
CT292	UK	QEHB	QEHB Contemporary	Screening	25/09/2014	1 3 0 3 15 - 2	1	3	0	3	15	-	2	22	R	33.73

CT284	UK	PEH	Guernsey	Clinical	29/09/2014	1 2 0 3 18 - 2	1	4	0	3	18	-	2	22	R	23.25
CT293	UK	QEHB	QEHB Contemporary	Clinical	16/10/2014	1 2 0 3 17 - 2	1	2				-	2	22	R	17.98
R1075	UK	HEFT	HEFT contemporary	Screening	21/10/2014	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	392.31
R1074	UK	HEFT	HEFT contemporary	Clinical	25/10/2014	1 3 0 3 17 - 2	1	3	0	3	17	-	2	22	R	251.79
R1080	UK	HEFT	HEFT contemporary	Screening	24/11/2014	1 3 0 3 17 - 2	1	3	0	3	17	-	2	22	R	100.03
R1081	UK	HEFT	HEFT contemporary	Screening	17/12/2014	1 3 0 3 17 - 2	1	3	0	3	17	-	2	22	R	119.81
R1126	UK	HEFT	HEFT contemporary	Clinical	30/12/2014	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	131.33
R1127	UK	HEFT	HEFT contemporary	Screening	02/01/2015	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	141.05
R1130	UK	HEFT	HEFT contemporary	Clinical	09/01/2015	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	146.41
R1128	UK	HEFT	HEFT contemporary	Clinical	15/01/2015	1 2 0 3 15 - 2	1	2	0	3	15	-	2	22	R	115.10
R1129	UK	HEFT	HEFT contemporary	Clinical	15/01/2015	1 2 0 3 15 - 2	1	2	0	3	15	-	2	22	R	115.00
CT294	UK	QEHB	QEHB Contemporary	Clinical	08/02/2015	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	23.67
CT295	UK	QEHB	QEHB Contemporary	Screening	12/03/2015	1 2 0 3 15 - 2	1	2	0	3	15	-	2	22	R	39.48
CT296	UK	QEHB	QEHB Contemporary	Screening	12/03/2015	1 2 0 3 14 - 2	1	3	0	3	14	-	2	22	R	23.39
CT297	UK	QEHB	QEHB Contemporary	Screening	14/03/2015	1 5 0 3 17 - 2	1	5	0	3	17	-	2	22	R	46.09
CT298	UK	QEHB	QEHB Contemporary	Clinical	16/03/2015	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	32.29
CT299	UK	QEHB	QEHB Contemporary	Clinical	25/03/2015	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	15.98
CT300	UK	QEHB	QEHB Contemporary	Clinical	25/03/2015	1 4 0 3 17 - 2	1	4	0	3	17	-	2	22	R	18.23
CT301	UK	QEHB	QEHB Contemporary	Clinical	31/03/2015	1 4 0 3 15 - 2	1	4	0	3	15	-	2	22	R	17.67
CT302	UK	QEHB	QEHB Contemporary	Screening	04/04/2015	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	30.79
CT303	UK	QEHB	QEHB Contemporary	Clinical	04/04/2015	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	27.80
CT304	UK	QEHB	QEHB Contemporary	Clinical	06/04/2015	1 5 0 3 17 - 2	1	5	0	3	17	-	2	22	R	51.02
CT305	UK	QEHB	QEHB Contemporary	Screening	08/04/2015	1 2 0 3 15 - 2	1	2	0	3	15	-	2	22	R	28.52
CT306	UK	QEHB	QEHB Contemporary	Screening	11/04/2015	1 4 0 3 15 - 2	1	4	0	3	15	-	2	22	R	26.65
CT307	UK	QEHB	QEHB Contemporary	Screening	14/04/2015	1 4 0 3 15 - 2	1	4	0	3	15	-	2	22	R	40.27
CT308	UK	QEHB	QEHB Contemporary	Screening	17/04/2015	1 4 0 3 17 4 2	1	2	0	3	17	4	2	22	R	16.90
CT309	UK	QEHB	QEHB Contemporary	Clinical	19/04/2015	1 5 0 3 17 - 2	1	5	0	3	17	-	2	22	R	25.83
CT282	UK	PEH	Guernsey	Clinical	18/05/2015	1 2 0 3 18 3 2	1	4	0	3	18	3	2	22	R	30.21
CT267	UK	PEH	Guernsey	Clinical	20/05/2015	1 4 0 3 14 - 2	1	4	0	3	14	-	2	22	R	17.91

CT272	UK	PEH	Guernsey	Clinical	21/05/2015	1 3 0 3 18 - 2	1	3	0	3	18	-	2	22	R	21.87
CT177	UK	QEHB	QEHB Contemporary	Unknown	-	1 2 0 3 17 - 2	1	2	0	3	17	-	2	22	R	15.31
CT195	UK	QEHB	QEHB Contemporary	Unknown	-	1 2 0 3 16 - 2	1	2	0	3	16	-	2	22	R	18.96
ECMM322	UK	HEFT	HEFT contemporary	Screening	02/04/2007	1 2 0 3 16 - 2	1	2	0	3	16	-	2	22	R	31.07
R267	UK	PEH	Guernsey	Unknown	2010-2011	1 4 0 3 16 - 2	1	4	0	3	16	-	2	22	R	18.65
R306	UK	PEH	Guernsey	Unknown	2010-2011	1 4 0 3 17 2 2	1	4	0	3	17	2	2	22	R	23.76
R307	UK	PEH	Guernsey	Unknown	2010-2011	1 3 0 3 17 2 2	1	3	0	3	17	2	2	22	R	33.59
R415	UK	PEH	Guernsey	Unknown	2010-2011	1 3 0 3 17 - 2	1	3	0	3	17	-	2	22	R	32.17
R005	UK	PEH	Guernsey	Unknown	2010-2011	1 4 0 3 17 - 2	1	4	0	3	17	-	2	22	R	45.08
R034	UK	PEH	Guernsey	Unknown	2010-2011	1 4 0 3 17 - 2	1	4	0	3	17	-	2	22	R	34.71
R200	UK	PEH	Guernsey	Unknown	2010-2011	1 4 0 3 18 - 2	1	4	0	3	18	-	2	22	R	27.92
R239	UK	PEH	Guernsey	Unknown	2010-2011	1 4 0 3 17 - 2	1	4	0	3	17	-	2	22	R	15.49
R240	UK	PEH	Guernsey	Unknown	2010-2011	1 4 0 3 17 - 2	1	4	0	3	17	-	2	22	R	23.39
ECMM341	UK	HEFT	HEFT contemporary	Screening	21/04/2007	1 2 0 3 16 - 2	1	2	0	3	16	-	2	22	R	66.65
CT124	UK	CHB	CHB Contemporary	Unknown	c1985	1 4 0 3 16 2 2	1	4	0	3	16	2	2	22	S	20.68
HK_w17	Hong Kong	HK1	Hong Kong	Clinical	2010	1 4 0 3 18 - 2	1	4	0	3	18	-	2	22	R	132.57
HK_w32	Hong Kong	HK1	Hong Kong	Clinical	2010	1 4 0 3 18 - 2	1	4	0	3	18	-	2	22	R	111.06
HK_w16	Hong Kong	HK4	Hong Kong	Clinical	2010	1 4 0 3 18 - 2	1	4	0	3	18	-	2	22	R	105.43
HK_w129	Hong Kong	HK1	Hong Kong	Clinical	2011	1 4 0 3 18 - 2	1	4	0	3	18	-	2	22	R	105.56
HK_N02	Hong Kong	HK2	Hong Kong	Screening	2014	1 4 0 3 18 - 2	1	4	0	3	18	-	2	22	R	100.52
HK_N03	Hong Kong	HK2	Hong Kong	Screening	2014	1 4 0 3 18 - 2	1	4	0	3	18	-	2	22	R	121.96
HK_N04	Hong Kong	HK2	Hong Kong	Screening	2014	1 4 0 3 18 - 2	1	4	0	3	18	-	2	22	R	94.55
HK_N05	Hong Kong	HK2	Hong Kong	Screening	2014	1 4 0 3 18 - 2	1	4	0	3	18	-	2	22	R	176.12
HK_N12	Hong Kong	HK1	Hong Kong	Screening	2014	1 4 0 3 17 - 2	1	4	0		17	-	2	22	R	138.29
HK_N13	Hong Kong	HK1	Hong Kong	Screening	2014	1 4 0 3 18 - 2	1	4	0	3	18	-	2	22	R	98.22
HK_N16	Hong Kong	HK1	Hong Kong	Screening	2014	1 4 0 3 18 - 2	1	4	0	3	18	-	2	22	R	95.30
HK_N17	Hong Kong	HK1	Hong Kong	Screening	2014	1 4 0 3 18 - 2	1	4	0	3	18	-	2	22	R	118.73
HK_N01	Hong Kong	HK2	Hong Kong	Screening	2014	1 4 0 3 18 - 2	1	4	0	3	18	-	2	22	R	124.44

$$\frac{\text{Number of colonies in region}}{\text{Volume of region in mls}} = \text{cfu/ml}$$

e.g. 31 + 31 colonies (in two opposing quadrants)
0.0015ml deposition volume

$$\frac{31 + 31}{0.0015} = 4.1\text{x}10^4 \text{ cfu/ml}$$

* NEEDS TO BE WORKED BACK UP TO ORIGINAL VOLUME

> Orignially had 20µl spot that was diluted in 2mls
= 20ul in 2000µl
= 1 in 100 dilution

> then took 500 µl of diluent and mixed with 3.5ml of water
= 500µl in 3500µl
= 1 in 8 dilution

Total dulution ratio = 1 in 800
= x cfu values from spiral plater by 800

4.1x10^4 x 800 = 3.3x10^7 cfu/ml

Figure 29: Example calculation of colony counts

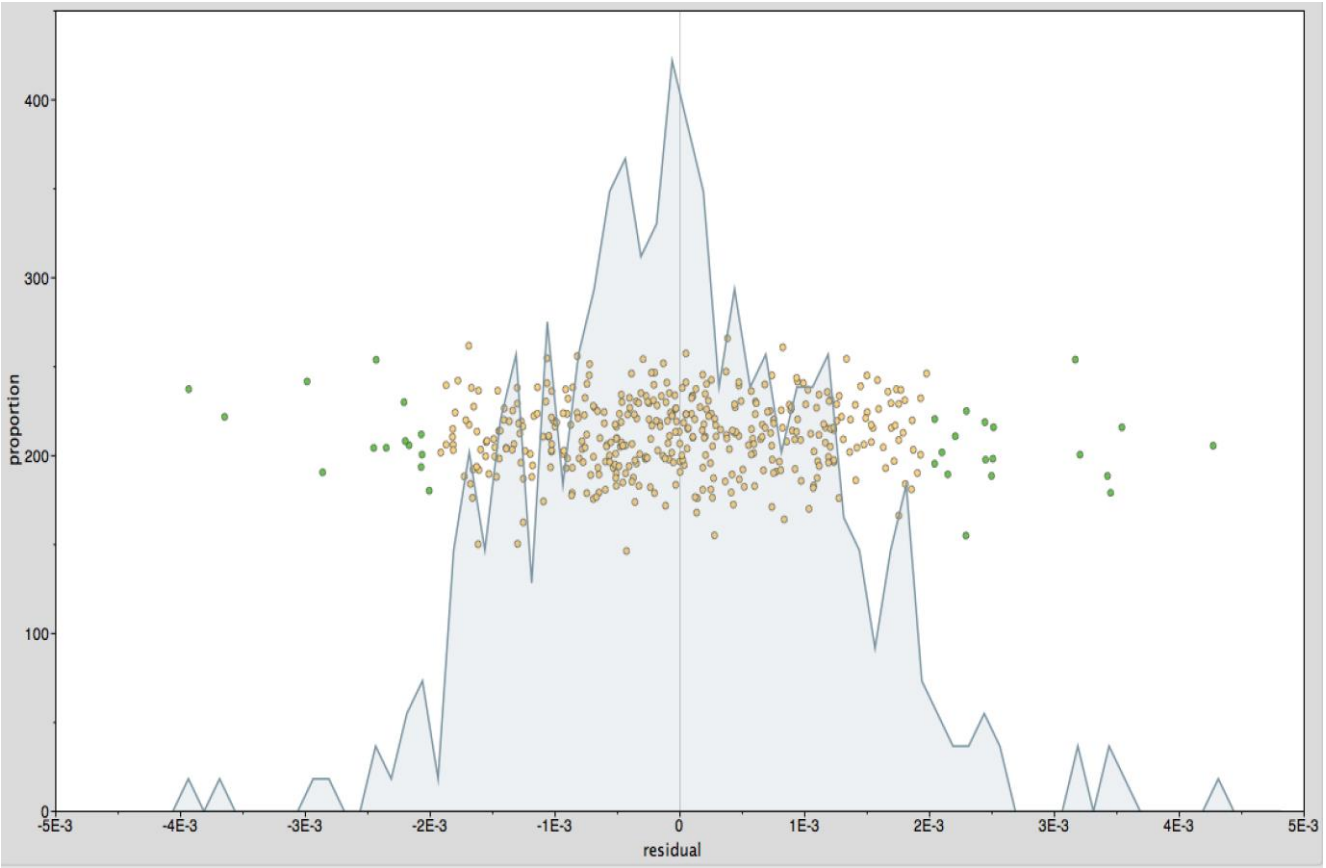

Figure 30: Scatter plot of residual value distance from the strict molecular clock line with a histogram of isolate proportion in grey. Residual values falling beyond +/- 2E-3 are coloured green.

6 Bibliography

Aanensen, D. M., Feil, E. J., Holden, M. T. G., Dordel, J., Yeats, C. A., Fedosejev, A., Goater, R., Castillo-Ramírez, S., Corander, J., Colijn, C., Chlebowicz, M. A., Schouls, L., Heck, M., Pluister, G., Ruimy, R., Kahlmeter, G., Åhman, J., Matuschek, E., Friedrich, A. W., Parkhill, J., Bentley, S. D., Spratt, B. G. & Grundmann, H. (2016). Whole-Genome Sequencing for Routine Pathogen Surveillance in Public Health: a Population Snapshot of Invasive *Staphylococcus aureus* in Europe. *mBio,* **7**.

Ahlstrom, C., Barkema, H. W., Stevenson, K., Zadoks, R. N., Biek, R., Kao, R., Trewby, H., Haupstein, D., Kelton, D. F., Fecteau, G., Labrecque, O., Keefe, G. P., Mckenna, S. L. B. & De Buck, J. (2015). Limitations of variable number of tandem repeat typing identified through whole genome sequencing of *Mycobacterium avium* subsp. paratuberculosis on a national and herd level. *BMC Genomics,* **16,** 161.

Ahn, S.-M., Kim, T.-H., Lee, S., Kim, D., Ghang, H., Kim, D.-S., Kim, B.-C., Kim, S.-Y., Kim, W.-Y., Kim, C., Park, D., Lee, Y. S., Kim, S., Reja, R., Jho, S., Kim, C. G., Cha, J.-Y., Kim, K.-H., Lee, B., Bhak, J. & Kim, S.-J. (2009). The first Korean genome sequence and analysis: Full genome sequencing for a socio-ethnic group. *Genome Research,* **19,** 1622-1629.

Aiello, A. E., Lowy, F. D., Wright, L. N. & Larson, E. L. (2006). Meticillin-resistant *Staphylococcus aureus* among US prisoners and military personnel: review and recommendations for future studies. *Lancet Infect.Dis.,* **6,** 335-341.

Alam, M. T., Read, T. D., Petit, R. A., Boyle-Vavra, S., Miller, L. G., Eells, S. J., Daum, R. S. & David, M. Z. (2015). Transmission and Microevolution of USA300 MRSA in U.S. Households: Evidence from Whole-Genome Sequencing. *mBio,* **6**.

Albrecht, N., Jatzwauk, L., Slickers, P., Ehricht, R. & Monecke, S. (2011). Clonal replacement of methicillin-resistant *Staphylococcus aureus* strains in a German university hospital. *PLoS One,* **6**.

Albrecht, V. S., Limbago, B. M., Moran, G. J., Krishnadasan, A., Gorwitz, R. J., Mcdougal, L. K. & Talan, D. A. (2015). *Staphylococcus aureus* Colonization and Strain Type at Various Body Sites among Patients with a Closed Abscess and Uninfected Controls at U.S. Emergency Departments. *Journal of Clinical Microbiology,* **53,** 3478-3484.

Aldeyab, M. A., Monnet, D. L., López-Lozano, J. M., Hughes, C. M., Scott, M. G., Kearney, M. P., Magee, F. A. & Mcelnay, J. C. (2008). Modelling the impact of antibiotic use and infection control practices on the incidence of hospital-acquired methicillin-resistant Staphylococcus aureus: a time-series analysis. *Journal of Antimicrobial Chemotherapy,* **62,** 593-600.

Aldred, K. J., Mcpherson, S. A., Turnbough, C. L., Kerns, R. J. & Osheroff, N. (2013). Topoisomerase IV-quinolone interactions are mediated through a water-metal ion

bridge: mechanistic basis of quinolone resistance. *Nucleic Acids Research,* **41,** 4628-4639.

Amorim, M. L., Faria, N. A., Oliveira, D. C., Vasconcelos, C., Cabeda, J. C., Mendes, A. C., Calado, E., Castro, A. P., Ramos, M. H., Amorim, J. M. & De, L. H. (2007). Changes in the clonal nature and antibiotic resistance profiles of methicillin-resistant *Staphylococcus aureus* isolates associated with spread of the EMRSA-15 clone in a tertiary care Portuguese hospital. *J.Clin.Microbiol.,* **45,** 2881-2888.

Anderson, V. E., Zaniewski, R. P., Kaczmarek, F. S., Gootz, T. D. & Osheroff, N. (1999). Quinolones Inhibit DNA Religation Mediated by *Staphylococcus aureus* Topoisomerase IV: changes in drug mechanism across evolutionary boundaries. *Journal of Biological Chemistry,* **274,** 35927-35932.

Andersson, M. I. & Macgowan, A. P. (2003). Development of the quinolones. *Journal of Antimicrobial Chemotherapy,* **51,** 1-11.

Andrews, J. M. (2001). Determination of minimum inhibitory concentrations. *Journal of Antimicrobial Chemotherapy,* **48,** 5-16.

Ashiru-Oredope, D., Sharland, M., Charani, E., Mcnulty, C. & Cooke, J. (2012). Improving the quality of antibiotic prescribing in the NHS by developing a new Antimicrobial Stewardship Programme: Start Smart—Then Focus. *Journal of Antimicrobial Chemotherapy,* **67,** i51-i63.

Atkinson, S. R., Paul, J., Sloan, E., Curtis, S. & Miller, R. (2009). The emergence of meticillin-resistant *Staphylococcus aureus* among injecting drug users. *Journal of Infection,* **58,** 339-345.

Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A. & Saunders, N. C. (1987). Intraspecific Phylogeography: The Mitochondrial DNA Bridge Between Population Genetics and Systematics. *Annual Review of Ecology and Systematics,* **18,** 489-522.

Baker, S., Duy, P. T., Nga, T. V. T., Dung, T. T. N., Phat, V. V., Chau, T. T., Turner, A. K., Farrar, J. & Boni, M. F. (2013). Fitness benefits in fluoroquinolone-resistant *Salmonella Typhi* in the absence of antimicrobial pressure. *eLife,* **2,** e01229.

Baldan, R., Rancoita, P. M., Di, S. C., Mazzotti, M., Cichero, P., Ossi, C., Biancardi, A., Nizzero, P., Saracco, A., Scarpellini, P. & Cirillo, D. M. (2015). Epidemic MRSA clone ST22-IV is more resistant to multiple host- and environment-related stresses compared with ST228-I. *Journal of Antimicrobial Chemotherapy,* **70,** 757-765.

Baldan, R., Testa, F., Lore, N. I., Bragonzi, A., Cichero, P., Ossi, C., Biancardi, A., Nizzero, P., Moro, M. & Cirillo, D. M. (2012). Factors contributing to epidemic MRSA clones replacement in a hospital setting. *PLoS.One.,* **7,** e43153.

Ball, P. (2000). Quinolone generations: natural history or natural selection? *Journal of Antimicrobial Chemotherapy,* **46,** 17-24.

Bao, R., Huang, L., Andrade, J., Tan, W., Kibbe, W. A., Jiang, H. & Feng, G. (2014). Review of Current Methods, Applications, and Data Management for the Bioinformatics Analysis of Whole Exome Sequencing. *Cancer Informatics,* **13,** 67-82.

Baranello, L., Levens, D., Gupta, A. & Kouzine, F. (2012). The importance of being supercoiled: How DNA mechanics regulate dynamic processes. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms,* **1819,** 632-638.

Bartels, M. D., Petersen, A., Worning, P., Nielsen, J. B., Larner-Svensson, H., Johansen, H. K., Andersen, L. P., Jarlov, J. O., Boye, K., Larsen, A. R. & Westh, H. (2014). Comparing whole-genome sequencing with Sanger sequencing for spa typing of methicillin-resistant Staphylococcus aureus. *J.Clin.Microbiol.,* **52,** 4305-4308.

Bassetti, S. & Battegay, M. (2004). *Staphylococcus aureus* Infections in Injection Drug Users: Risk Factors and Prevention Strategies. *Infection,* **32,** 163-169.

Beard-Pegler, M. A., Stubbs, E. & Vickery, A. M. (1988). Observations on the resistance to drying of staphylococcal strains. *Journal of Medical Microbiology,* **26,** 251-255.

Benner, J. E. & Kayser, F. H. (1968). Growing clinical significance of methicillin resistant *Staphylococcus aureus. The Lancet,* **292,** 741-744.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Cheetham, R. K., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray, S. J., Irving, L. J., Karbelashvili, M. S., Kirk, S. M., Li, H., Liu, X., Maisinger, K. S., Murray, L. J., Obradovic, B., Ost, T., Parkinson, M. L., Pratt, M. R., Rasolonjatovo, I. M. J., Reed, M. T., Rigatti, R., Rodighiero, C., Ross, M. T., Sabot, A., Sankar, S. V., Scally, A., Schroth, G. P., Smith, M. E., Smith, V. P., Spiridou, A., Torrance, P. E., Tzonev, S. S., Vermaas, E. H., Walter, K., Wu, X., Zhang, L., Alam, M. D., Anastasi, C., Aniebo, I. C., Bailey, D. M. D., Bancarz, I. R., Banerjee, S., Barbour, S. G., Baybayan, P. A., Benoit, V. A., Benson, K. F., Bevis, C., Black, P. J., Boodhun, A., Brennan, J. S., Bridgham, J. A., Brown, R. C., Brown, A. A., Buermann, D. H., Bundu, A. A., Burrows, J. C., Carter, N. P., Castillo, N., Catenazzi, M. C. E., Chang, S., Cooley, R. N., Crake, N. R., Dada, O. O., Diakoumakos, K. D., Dominguez-Fernandez, B., Earnshaw, D. J., Egbujor, U. C., Elmore, D. W., Etchin, S. S., Ewan, M. R., Fedurco, M., Fraser, L. J., Fajardo, K. V. F., Furey, W. S., George, D., Gietzen, K. J., Goddard, C. P., Golda, G. S., Granieri, P. A., Green, D. E., Gustafson, D. L., Hansen, N. F., Harnish, K., Haudenschild, C. D., Heyer, N. I., Hims, M. M., Ho, J. T., Horgan, A. M., et al. (2008). Accurate Whole Human Genome Sequencing using Reversible Terminator Chemistry. *Nature,* **456,** 53-59.

Bischoff, M., Dunman, P., Kormanec, J., Macapagal, D., Murphy, E., Mounts, W., Berger-Bächi, B. & Projan, S. (2004). Microarray-Based Analysis of the *Staphylococcus aureus* σ(B) Regulon. *Journal of Bacteriology,* **186,** 4085-4099.

Bletz, S., Mellmann, A., Rothganger, J. & Harmsen, D. (2015). Ensuring backwards compatibility: traditional genotyping efforts in the era of whole genome sequencing. *Clin.Microbiol.Infect.,* **21,** 347-4.

Bode, L. G., Kluytmans, J. A., Wertheim, H. F., Bogaers, D., Vandenbroucke-Grauls, C. M., Roosendaal, R., Troelstra, A., Box, A. T., Voss, A., Van, D. T. I., Van, B. A., Verbrugh, H. A. & Vos, M. C. (2010). Preventing surgical-site infections in nasal carriers of *Staphylococcus aureus*. *New England Journal of Medicine,* **362,** 9-17.

Bolger, A. M., Lohse, M. & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* **30,** 2114-2120.

Boyce, J. M. (2007). Environmental contamination makes an important contribution to hospital infection. *Journal of Hospital Infection,* **65, Supplement 2,** 50-54.

Boyce, J. M. (2016). Modern technologies for improving cleaning and disinfection of environmental surfaces in hospitals. *Antimicrobial Resistance & Infection Control,* **5,** 10.

Boyce, J. M., White, R. L. & Spruill, E. Y. (1983). Impact of Methicillin-Resistant *Staphylococcus aureus* on the Incidence of Nosocomial Staphylococcal Infections. *Journal of Infectious Diseases,* **148,** 763.

Brennan, L., Lilliebridge, R. A., Cheng, A. C., Giffard, P. M., Currie, B. J. & Tong, S. Y. C. (2013). Community-associated meticillin-resistant *Staphylococcus aureus* carriage in hospitalized patients in tropical northern Australia. *Journal of Hospital Infection,* **83,** 205-211.

Cairns, M. D., Preston, M. D., Lawley, T. D., Clark, T. G., Stabler, R. A. & Wren, B. W. (2015). Genomic Epidemiology of a Protracted Hospital Outbreak Caused by a Toxin A-Negative C*lostridium difficile* Sublineage PCR Ribotype 017 Strain in London, England. *Journal of Clinical Microbiology,* **53,** 3141-3147.

Calfee, D. P., Salgado, C. D., Milstone, A. M., Harris, A. D., Kuhar, D. T., Moody, J., Aureden, K., Huang, S. S., Maragakis, L. L. & Yokoe, D. S. (2014). Strategies to Prevent Methicillin-Resistant *Staphylococcus aureus* Transmission and Infection in Acute Care Hospitals: 2014 Update. *Infection Control and Hospital Epidemiology,* **35,** 772-796.

Carling, P. C. & Bartley, J. M. (2010). Evaluating hygienic cleaning in health care settings: What you do not know can harm your patients. *American Journal of Infection Control,* **38,** S41-S50.

Castillo-Ramírez, S., Corander, J., Marttinen, P., Aldeljawi, M., Hanage, W. P., Westh, H., Boye, K., Gulay, Z., Bentley, S. D., Parkhill, J., Holden, M. T. & Feil, E. J. (2012). Phylogeographic variation in recombination rates within a global clone of methicillin-resistant *Staphylococcus aureus*. *Genome Biology,* **13,** R126-R126.

Chaibenjawong, P. & Foster, S. J. (2010). Desiccation tolerance in *Staphylococcus aureus*. *Archives of Microbiology,* **193,** 125-135.

Champion, A. E., Goodwin, T. A., Brolinson, P. G., Werre, S. R., Prater, M. R. & Inzana, T. J. (2014). Prevalence and characterization of methicillin-resistant *Staphylococcus aureus* isolates from healthy university student athletes. *Annals of Clinical Microbiology and Antimicrobials,* **13,** 33-33.

Charbonneau, P., Parienti, J.-J., Thibon, P., Ramakers, M., Daubin, C., Du Cheyron, D., Lebouvier, G., Le Coutour, X., Leclercq, R. & Group, F. F. F. S. (2006). Fluoroquinolone Use and Methicillin-Resistant *Staphylococcus aureus* Isolation Rates in Hospitalized Patients: A Quasi Experimental Study. *Clinical Infectious Diseases,* **42,** 778-784.

Chemaly, R. F., Simmons, S., Dale, C., Ghantoji, S. S., Rodriguez, M., Gubb, J., Stachowiak, J. & Stibich, M. (2014). The role of the healthcare environment in the spread of multidrug-resistant organisms: update on current best practices for containment. *Therapeutic Advances in Infectious Disease,* **2,** 79-90.

Chen, C. J. & Huang, Y. C. (2014). New epidemiology of *Staphylococcus aureus* infection in Asia. *Clinical Microbiology and Infection,* **20,** 605-623.

Chen, Y., Liu, Z., Duo, L., Xiong, J., Gong, Y., Yang, J., Wang, Z., Wu, X., Lu, Z., Meng, X., Zhao, J., Zhang, C., Wang, F., Zhang, Y., Zhang, M. & Han, L. (2014). Characterization of *Staphylococcus aureus* from Distinct Geographic Locations in China: An Increasing Prevalence of *spa*-t030 and SCC*mec* Type III. *PLoS ONE,* **9,** e96255.

Chongtrakool, P., Ito, T., Ma, X. X., Kondo, Y., Trakulsomboon, S., Tiensasitorn, C., Jamklang, M., Chavalit, T., Song, J.-H. & Hiramatsu, K. (2006). Staphylococcal Cassette Chromosome mec (SCCmec) Typing of Methicillin-Resistant *Staphylococcus aureus* Strains Isolated in 11 Asian Countries: a Proposal for a New Nomenclature for SCCmec Elements. *Antimicrobial Agents and Chemotherapy,* **50,** 1001-1012.

Clauditz, A., Resch, A., Wieland, K.-P., Peschel, A. & Götz, F. (2006). Staphyloxanthin Plays a Role in the Fitness of *Staphylococcus aureus* and Its Ability To Cope with Oxidative Stress. *Infection and Immunity,* **74,** 4950-4953.

Clements, M. O. & Foster, S. J. (1999). Stress resistance in *Staphylococcus aureus*. *Trends in Microbiology,* **7,** 458-462.

Coia, J. E., Duckworth, G. J., Edwards, D. I., Farrington, M., Fry, C., Humphreys, H., Mallaghan, C. & Tucker, D. R. Guidelines for the control and prevention of meticillin-resistant *Staphylococcus aureus* (MRSA) in healthcare facilities. *Journal of Hospital Infection,* **63,** S1-S44.

Cooke, F. J., Gkrania-Klotsas, E., Howard, J. C., Stone, M., Kearns, A. M., Ganner, M., Carmichael, A. J. & Brown, N. M. (2010). Clinical, molecular and epidemiological description of a cluster of community-associated methicillin-resistant *Staphylococcus aureus* isolates from injecting drug users with bacteraemia. *Clinical Microbiology and Infection,* **16,** 921-926.

Cooke, F. J., Howard, J. C., Hugh-Jones, C. & Brown, N. M. (2008). Meticillin-resistant *Staphylococcus aureus* in the community: homeless are also at risk. *Journal of Hospital Infection,* **68,** 186-188.

Cooke, J., Stephens, P., Ashiru-Oredope, D., Charani, E., Dryden, M., Fry, C., Hand, K., Holmes, A., Howard, P., Johnson, A. P., Livermore, D. M., Mansell, P., Mcnulty, C. a. M., Wellsteed, S., Hopkins, S. & Sharland, M. (2015). Longitudinal trends and cross-sectional

analysis of English national hospital antibacterial use over 5 years (2008–13): working towards hospital prescribing quality measures. *Journal of Antimicrobial Chemotherapy,* **70,** 279-285.

Coombs, G. W., Nimmo, G. R., Bell, J. M., Huygens, F., O'brien, F. G., Malkowski, M. J., Pearson, J. C., Stephens, A. J., Giffard, P. M. & Resistance, T. a. G. F. A. (2004). Genetic Diversity among Community Methicillin-Resistant *Staphylococcus aureus* Strains Causing Outpatient Infections in Australia. *Journal of Clinical Microbiology,* **42,** 4735-4743.

Coombs, G. W., Nimmo, G. R., Daly, D. A., Le, T. T., Pearson, J. C., Tan, H. L., Robinson, J. O., Collignon, P. J., Mclaws, M. L. & Turnidge, J. D. (2014). Australian *Staphylococcus aureus* Sepsis Outcome Programme annual report, 2013. *Commun.Dis.Intell.Q.Rep.,* **38,** E309-E319.

Coombs, G. W., Nimmo, G. R., Pearson, J. C., Collignon, P. J., Bell, J. M., Mclaws, M. L., Christiansen, K. J. & Turnidge, J. D. (2011). Australian Group on Antimicrobial Resistance Hospital-onset *Staphylococcus aureus* Surveillance Programme annual report, 2011. *Commincable diesases intelligence quarterly report* **30,** E210-E218.

Cosgrove, K., Coutts, G., Jonsson, I.-M., Tarkowski, A., Kokai-Kun, J. F., Mond, J. J. & Foster, S. J. (2007). Catalase (KatA) and Alkyl Hydroperoxide Reductase (AhpC) Have Compensatory Roles in Peroxide Stress Resistance and Are Required for Survival, Persistence, and Nasal Colonization in *Staphylococcus aureus*. *Journal of Bacteriology,* **189,** 1025-1035.

Cosgrove, S. E., Qi, Y., Kaye, K. S., Harbarth, S., Karchmer, A. W. & Carmeli, Y. (2005). The Impact of Methicillin Resistance in *Staphylococcus aureus* Bacteremia on Patient Outcomes: Mortality, Length of Stay, and Hospital Charges. *Infection Control and Hospital Epidemiology,* **26,** 166-174.

Cosgrove, S. E., Sakoulas, G., Perencevich, E. N., Schwaber, M. J., Karchmer, A. W. & Carmeli, Y. (2003). Comparison of Mortality Associated with Methicillin-Resistant and Methicillin-Susceptible *Staphylococcus aureus* Bacteremia: A Meta-analysis. *Clinical Infectious Diseases,* **36,** 53-59.

Croucher, N. J., Harris, S. R., Grad, Y. H. & Hanage, W. P. (2013). Bacterial genomes in epidemiology—present and future. *Philosophical Transactions of the Royal Society B: Biological Sciences,* **368,** 20120202.

Croucher, N. J., Page, A. J., Connor, T. R., Delaney, A. J., Keane, J. A., Bentley, S. D., Parkhill, J. & Harris, S. R. (2015). Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.,* **43,** e15.

Crozat, E., Philippe, N., Lenski, R. E., Geiselmann, J. & Schneider, D. (2005). Long-Term Experimental Evolution in *Escherichia coli.* XII. DNA Topology as a Key Target of Selection. *Genetics,* **169,** 523-532.

Crump, B., Wise, R. & Dent, J. (1983). Pharmacokinetics and tissue penetration of ciprofloxacin. *Antimicrobial Agents and Chemotherapy,* **24,** 784-786.

Dailey, L., Coombs, G. W., O'brien, F. G., Pearman, J. W., Christiansen, K., Grubb, W. B. & Riley, T. V. (2005). Methicillin-resistant *Staphylococcus aureus*, Western Australia. *Emerging Infectious Diseases,* **11,** 1584-1590.

David, M. Z., Cadilla, A., Boyle-Vavra, S. & Daum, R. S. (2014). Replacement of HA-MRSA by CA-MRSA infections at an academic medical center in the midwestern United States, 2004-5 to 2008. *PLoS.One.,* **9,** e92760.

Davis, R. & Bryson, H. M. (1994). Levofloxacin. *Drugs,* **47,** 677-700.

Deshpande, A., Pasupuleti, V., Thota, P., Pant, C., Rolston, D. D. K., Hernandez, A. V., Donskey, C. J. & Fraser, T. G. (2015). Risk Factors for Recurrent *Clostridium difficile* Infection: A Systematic Review and Meta-Analysis. *Infection Control & Hospital Epidemiology,* **36,** 452-460.

Deurenberg, R. H. & Stobberingh, E. E. (2008). The evolution of *Staphylococcus aureus*. *Infection, Genetics and Evolution* **8,** 747-763.

Dhawan, B., Rao, C., Udo, E. E., Gadepalli, R., Vishnubhatla, S. & Kapil, A. (2015). Dissemination of methicillin-resistant *Staphylococcus aureus* SCCmec type IV and SCCmec type V epidemic clones in a tertiary hospital: challenge to infection control. *Epidemiol.Infect.,* **143,** 343-353.

Didelot, X., Eyre, D. W., Cule, M., Ip, C. L., Ansari, M. A., Griffiths, D., Vaughan, A., O'connor, L., Golubchik, T., Batty, E. M., Piazza, P., Wilson, D. J., Bowden, R., Donnelly, P. J., Dingle, K. E., Wilcox, M., Walker, A. S., Crook, D. W., Te, A. P. & Harding, R. M. (2012). Microevolutionary analysis of *Clostridium difficile* genomes to investigate transmission. *Genome Biol.,* **13,** R118.

Didelot, X., Lawson, D., Darling, A. & Falush, D. (2010). Inference of Homologous Recombination in Bacteria Using Whole-Genome Sequences. *Genetics,* **186,** 1435-1449.

Didelot, X. & Maiden, M. C. J. (2010). Impact of recombination on bacterial evolution. *Trends in microbiology,* **18,** 315-322.

Diekema, D. J., Pfaller, M. A., Schmitz, F. J., Smayevsky, J., Bell, J., Jones, R. N. & Beach, M. (2001). Survey of Infections Due to Staphylococcus Species: Frequency of Occurrence and Antimicrobial Susceptibility of Isolates Collected in the United States, Canada, Latin America, Europe, and the Western Pacific Region for the SENTRY Antimicrobial Surveillance Program, 1997–1999. *Clinical Infectious Diseases,* **32,** S114-S132.

Domon, H., Uehara, Y., Oda, M., Seo, H., Kubota, N. & Terao, Y. (2016). Poor survival of Methicillin‐resistant *Staphylococcus aureus* on inanimate objects in the public spaces. *MicrobiologyOpen,* **5,** 39-46.

Driebe, E. M., Sahl, J. W., Roe, C., Bowers, J. R., Schupp, J. M., Gillece, J. D., Kelley, E., Price, L. B., Pearson, T. R., Hepp, C. M., Brzoska, P. M., Cummings, C. A., Furtado, M. R., Andersen, P. S., Stegger, M., Engelthaler, D. M. & Keim, P. S. (2015). Using Whole Genome Analysis

to Examine Recombination across Diverse Sequence Types of *Staphylococcus aureus*. *PLoS.One.,* **10,** e0130955.

Drlica, K. (1990). Bacterial topoisomerases and the control of DNA supercoiling. *Trends in Genetics,* **6,** 433-437.

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

Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. (2012). Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution,* **29,** 1969-1973.

Duckworth, G. J. & Jordens, J. Z. (1990). Adherence and survival properties of an epidemic methicillin-resistant strain of *Staphylococcus aureus* compared with those of methicillin-sensitive strains. *Journal of Medical Microbiology,* **32,** 195-200.

Duckworth, G. J., Lothian, J. L. E. & Williams, J. D. (1988). Methicillin-resistant *Staphylococcus aureus*: report of an outbreak in a London teaching hospital. *Journal of Hospital Infection,* **11,** 1-15.

Duerden, B., Fry, C., Johnson, A. P. & Wilcox, M. H. (2015). The Control of Methicillin-Resistant *Staphylococcus aureus* Blood Stream Infections in England. *Open Forum Infectious Diseases,* **2,** ofv035.

Edgeworth, J. D., Yadegarfar, G., Pathak, S., Batra, R., Cockfield, J. D., Wyncoll, D., Beale, R. & Lindsay, J. A. (2007). An outbreak in an intensive care unit of a strain of methicillin-resistant *Staphylococcus aureus* sequence type 239 associated with an increased rate of vascular access device-related bacteremia. *Clin.Infect.Dis.,* **44,** 493-501.

Edmond, M. B. & Wenzel, R. P. (2013). Screening inpatients for MRSA--case closed. *N.Engl.J.Med.,* **368,** 2314-2315.

Ekblom, R. & Wolf, J. B. W. (2014). A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications,* **7,** 1026-1042.

Elgohari, S., Wilson, J., Saei, A., Sheridan, E. A. & Lamagni, T. (2016). Impact of national policies on the microbial aetiology of surgical site infections in acute NHS hospitals in England: analysis of trends between 2000 and 2013 using multi-centre prospective cohort data. *Epidemiology & Infection,* **28,** 1-13.

Ellington, M. J., Hope, R., Livermore, D. M., Kearns, A. M., Henderson, K., Cookson, B. D., Pearson, A. & Johnson, A. P. (2010). Decline of EMRSA-16 amongst methicillin-resistant *Staphylococcus aureus* causing bacteraemias in the UK between 2001 and 2007. *J.Antimicrob.Chemother.,* **65,** 446-448.

Enright, M. C. (2003). The evolution of a resistant pathogen--the case of MRSA. *Curr.Opin.Pharmacol.,* **3,** 474-479.

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

European Comittee on Antimicrobial Susceptibility Testing (Eucast) (2016). Breakpoint tables for interpretation of MICs and zone diameters.

Eyre, D. W., Cule, M. L., Wilson, D. J., Griffiths, D., Vaughan, A., O'connor, L., Ip, C. L. C., Golubchik, T., Batty, E. M., Finney, J. M., Wyllie, D. H., Didelot, X., Piazza, P., Bowden, R., Dingle, K. E., Harding, R. M., Crook, D. W., Wilcox, M. H., Peto, T. E. A. & Walker, A. S. (2013). Diverse Sources of *C. difficile* Infection Identified on Whole-Genome Sequencing. *New England Journal of Medicine,* **369,** 1195-1205.

Eyre, D. W., Golubchik, T., Gordon, N. C., Bowden, R., Piazza, P., Batty, E. M., Ip, C. L., Wilson, D. J., Didelot, X., O'connor, L., Lay, R., Buck, D., Kearns, A. M., Shaw, A., Paul, J., Wilcox, M. H., Donnelly, P. J., Peto, T. E., Walker, A. S. & Crook, D. W. (2012). A pilot study of rapid benchtop sequencing of *Staphylococcus aureus* and *Clostridium difficile* for outbreak detection and surveillance. *BMJ Open.,* **2**.

Feil, E. J., Nickerson, E. K., Chantratita, N., Wuthiekanun, V., Srisomang, P., Cousins, R., Pan, W., Zhang, G., Xu, B., Day, N. P. & Peacock, S. J. (2008). Rapid detection of the pandemic methicillin-resistant *Staphylococcus aureus* clone ST 239, a dominant strain in Asian hospitals. *J.Clin.Microbiol.,* **46,** 1520-1522.

Feil, E. J. & Spratt, B. G. (2001). Recombination and the Population Structures of Bacterial Pathogens. *Annual Review of Microbiology,* **55,** 561-590.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution,* **17,** 368-376.

Felsenstein, J. (1985). Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution,* **39,** 783-791.

Finch, R., Whitby, M., Craddock, C., Holliday, A., Martin, J. & Pilkington, R. 1986. Clinical Evaluation of Treatment with Ciprofloxacin. *In:* NEU, H. C. & REEVES, D. S. (eds.) *Ciprofloxacin: Microbiology — Pharmacokinetics — Clinical Experience.* Wiesbaden: Vieweg+Teubner Verlag.

Fitzgerald, J. R. (2014). Evolution of *Staphylococcus aureus* during human colonization and infection. *Infection, Genetics and Evolution,* **21,** 542-547.

Fitzpatrick, M. A., Ozer, E. A. & Hauser, A. R. (2016). Utility of Whole-Genome Sequencing in Characterizing Acinetobacter Epidemiology and Analyzing Hospital Outbreaks. *Journal of Clinical Microbiology,* **54,** 593-612.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M. & Al, E. (1995). Whole-genome random sequencing and assembly o*f Haemophilus influenzae* Rd. *Science,* **269,** 496-512.

Ford, C. B., Lin, P. L., Chase, M. R., Shah, R. R., Iartchouk, O., Galagan, J., Mohaideen, N., Ioerger, T. R., Sacchettini, J. C., Lipsitch, M., Flynn, J. L. & Fortune, S. M. (2011). Use of whole genome sequencing to estimate the mutation rate of *Mycobacterium tuberculosis* during latent infection. *Nat Genet,* **43,** 482-486.

Fraud, S., Maillard, J. Y. & Russell, A. D. (2001). Comparison of the mycobactericidal activity of ortho- phthalaldehyde, glutaraldehyde and other dialdehydes by a quantitative suspension test. *Journal of Hospital Infection,* **48,** 214-221.

Frénay, H. M. E., Bunschoten, A. E., Schouls, L. M., Van Leeuwen, W. J., Vandenbroucke-Grauls, C. M. J. E., Verhoef, J. & Mooi, F. R. (1996). Molecular typing of methicillin-resistantStaphylococcus aureus on the basis of protein A gene polymorphism. *European Journal of Clinical Microbiology and Infectious Diseases,* **15,** 60-64.

Gardy, J., Loman, N. J. & Rambaut, A. (2015). Real-time digital pathogen surveillance — the time is now. *Genome Biology,* **16,** 155.

Gardy, J. L., Johnston, J. C., Ho Sui, S. J., Cook, V. J., Shah, L., Brodkin, E., Rempel, S., Moore, R., Zhao, Y., Holt, R., Varhol, R., Birol, I., Lem, M., Sharma, M. K., Elwood, K., Jones, S. J., Brinkman, F. S., Brunham, R. C. & Tang, P. (2011). Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *New England Journal of Medicine,* **364,** 730-739.

Garvey, M. I., Pichon, B., Bradley, C. W., Moiemen, N. S., Oppenheim, B. & Kearns, A. M. (2016). Improved understanding of an outbreak of meticillin-resistant *Staphylococcus aureus* in a regional burns centre via whole-genome sequencing. *Journal of Hospital Infection,* **94,** 401-404.

George, J. & Morrissey, I. (1997). The bactericidal activity of levofloxacin compared with ofloxacin, D-ofloxacin, ciprofloxacin, sparfloxacin and cefotaxime against *Streptococcus pneumoniae*. *Journal of Antimicrobial Chemotherapy,* **39,** 719-723.

Golding, G., Campbell, J., Spreitzer, D. & Chui, L. 2015. Pulsed-Field Gel Electrophoresis of *Staphylococcus aureus*. *In:* JORDAN, K. & DALMASSO, M. (eds.) *Pulse Field Gel Electrophoresis.* Springer New York.

Golubchik, T., Batty, E. M., Miller, R. R., Farr, H., Young, B. C., Larner-Svensson, H., Fung, R., Godwin, H., Knox, K., Votintseva, A., Everitt, R. G., Street, T., Cule, M., Ip, C. L. C., Didelot, X., Peto, T. E. A., Harding, R. M., Wilson, D. J., Crook, D. W. & Bowden, R. (2013). Within-Host Evolution of *Staphylococcus aureus* during Asymptomatic Carriage. *PLoS ONE,* **8,** e61319.

Goudarzi, M., Goudarzi, H., Sá Figueiredo, A. M., Udo, E. E., Fazeli, M., Asadzadeh, M. & Seyedjavadi, S. S. (2016). Molecular Characterization of Methicillin Resistant *Staphylococcus aureus* Strains Isolated from Intensive Care Units in Iran: ST22-SCC*mec* IV/t790 Emerges as the Major Clone. *PLoS ONE,* **11,** e0155529.

Gray, R. R., Tatem, A. J., Johnson, J. A., Alekseyenko, A. V., Pybus, O. G., Suchard, M. A. & Salemi, M. (2011). Testing Spatiotemporal Hypothesis of Bacterial Evolution Using

Methicillin-Resistant *Staphylococcus aureus* ST239 Genome-wide Data within a Bayesian Framework. *Molecular Biology and Evolution,* **28,** 1593-1603.

Griggs, D. J., Marona, H. & Piddock, L. J. (2003). Selection of moxifloxacin-resistant *Staphylococcus aureus* compared with five other fluoroquinolones. *J.Antimicrob.Chemother.,* **51,** 1403-1407.

Groves, M. D., Crouch, B., Coombs, G. W., Jordan, D., Pang, S., Barton, M. D., Giffard, P., Abraham, S. & Trott, D. J. (2016). Molecular Epidemiology of Methicillin-Resistant *Staphylococcus aureus* Isolated from Australian Veterinarians. *PLoS ONE,* **11,** e0146034.

Grundmann, H., Aanensen, D. M., Van Den Wijngaard, C. C., Spratt, B. G., Harmsen, D., Friedrich, A. W. & The European Staphylococcal Reference Laboratory Working, G. (2010). Geographic Distribution of *Staphylococcus aureus* Causing Invasive Infections in Europe: A Molecular-Epidemiological Analysis. *PLoS Medicine,* **7,** e1000215.

Grundmann, H., Aires-De-Sousa, M., Boyce, J. & Tiemersma, E. (2006). Emergence and resurgence of meticillin-resistant *Staphylococcus aureus* as a public-health threat. *The Lancet,* **368,** 874-885.

Grundmann, H., Schouls Lm Fau - Aanensen, D. M., Aanensen Dm Fau - Pluister, G. N., Pluister Gn Fau - Tami, A., Tami a Fau - Chlebowicz, M., Chlebowicz M Fau - Glasner, C., Glasner C Fau - Sabat, A. J., Sabat Aj Fau - Weist, K., Weist K Fau - Heuer, O., Heuer O Fau - Friedrich, A. W. & Friedrich, A. W. (2014). The dynamic changes of dominant clones of *Staphylococcus aureus* causing bloodstream infections in the European region: results of a second structured survey. LID - 20987 [pii].

Guzmán-Blanco, M., Mejía, C., Isturiz, R., Alvarez, C., Bavestrello, L., Gotuzzo, E., Labarca, J., Luna, C. M., Rodríguez-Noriega, E., Salles, M. J. C., Zurita, J. & Seas, C. (2009). Epidemiology of meticillin-resistant *Staphylococcus aureus* (MRSA) in Latin America. *International Journal of Antimicrobial Agents,* **34,** 304-308.

Han, J., Wang, Y., Sahin, O., Shen, Z., Guo, B., Shen, J. & Zhang, Q. (2012). A Fluoroquinolone Resistance Associated Mutation in *gyrA* Affects DNA Supercoiling in *Campylobacter jejuni*. *Frontiers in Cellular and Infection Microbiology,* **2,** 21.

Hardy, K. J., Oppenheim, B. A., Gossain, S., Gao, F. & Hawkey, P. M. (2006a). A study of the relationship between environmental contamination with methicillin-resistant *Staphylococcus aureus* (MRSA) and patients' acquisition of MRSA. *Infect.Control Hosp.Epidemiol.,* **27,** 127-132.

Hardy, K. J., Oppenheim, B. A., Gossain, S., Gao, F. & Hawkey, P. M. (2006b). Use of variations in staphylococcal interspersed repeat units for molecular typing of methicillin-resistant *Staphylococcus aureus* strains. *J.Clin.Microbiol.,* **44,** 271-273.

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

Harris, S. R., Cartwright, E. J., Torok, M. E., Holden, M. T., Brown, N. M., Ogilvy-Stuart, A. L., Ellington, M. J., Quail, M. A., Bentley, S. D., Parkhill, J. & Peacock, S. J. (2013). Whole-genome sequencing for analysis of an outbreak of meticillin-resistant *Staphylococcus aureus*: a descriptive study. *Lancet Infect.Dis.,* **13,** 130-136.

Harris, S. R., Feil, E. J., Holden, M. T., Quail, M. A., Nickerson, E. K., Chantratita, N., Gardete, S., Tavares, A., Day, N., Lindsay, J. A., Edgeworth, J. D., De, L. H., Parkhill, J., Peacock, S. J. & Bentley, S. D. (2010). Evolution of MRSA during hospital transmission and intercontinental spread. *Science,* **327,** 469-474.

Harrison, E. M., Weinert, L. A., Holden, M. T., Welch, J. J., Wilson, K., Morgan, F. J., Harris, S. R., Loeffler, A., Boag, A. K., Peacock, S. J., Paterson, G. K., Waller, A. S., Parkhill, J. & Holmes, M. A. (2014). A shared population of epidemic methicillin-resistant *Staphylococcus aureus* 15 circulates in humans and companion animals. *MBio.,* **5,** e00985-13.

Hawkey, P. M. (1997). Quinolones in sweat and quinolone resistance. *Lancet,* **349,** 148-149.

Hirai, Y. (1991). Survival of bacteria under dry conditions; from a viewpoint of nosocomial infection. *Journal of Hospital Infection,* **19,** 191-200.

Ho, S. Y. W. & Duchêne, S. (2014). Molecular-clock methods for estimating evolutionary rates and timescales. *Molecular Ecology,* **23,** 5947-5965.

Holden, M. T., Hsu, L. Y., Kurt, K., Weinert, L. A., Mather, A. E., Harris, S. R., Strommenger, B., Layer, F., Witte, W., De, L. H., Skov, R., Westh, H., Zemlickova, H., Coombs, G., Kearns, A. M., Hill, R. L., Edgeworth, J., Gould, I., Gant, V., Cooke, J., Edwards, G. F., Mcadam, P. R., Templeton, K. E., Mccann, A., Zhou, Z., Castillo-Ramirez, S., Feil, E. J., Hudson, L. O., Enright, M. C., Balloux, F., Aanensen, D. M., Spratt, B. G., Fitzgerald, J. R., Parkhill, J., Achtman, M., Bentley, S. D. & Nubel, U. (2013). A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant *Staphylococcus aureus* pandemic. *Genome Res.,* **23,** 653-664.

Holden, M. T., Lindsay, J. A., Corton, C., Quail, M. A., Cockfield, J. D., Pathak, S., Batra, R., Parkhill, J., Bentley, S. D. & Edgeworth, J. D. (2010). Genome sequence of a recently emerged, highly transmissible, multi-antibiotic- and antiseptic-resistant variant of methicillin-resistant *Staphylococcus aureus*, sequence type 239 (TW). *J.Bacteriol.,* **192,** 888-892.

Holt, D. C., Holden, M. T. G., Tong, S. Y. C., Castillo-Ramirez, S., Clarke, L., Quail, M. A., Currie, B. J., Parkhill, J., Bentley, S. D., Feil, E. J. & Giffard, P. M. (2011). A Very Early-Branching *Staphylococcus aureus* Lineage Lacking the Carotenoid Pigment Staphyloxanthin. *Genome Biology and Evolution,* **3,** 881-895.

Hooper, D. C. (1999). Mechanisms of fluoroquinolone resistance. *Drug Resistance Updates,* **2,** 38-55.

Hooper, D. C. (2000). Mechanisms of Action and Resistance of Older and Newer Fluoroquinolones. *Clinical Infectious Diseases,* **31,** S24-S28.

Hooper, D. C. (2001). Emerging mechanisms of fluoroquinolone resistance. *Emerging Infectious Diseases,* **7,** 337-341.

Horii, T., Suzuki, Y., Monji, A., Morita, M., Muramatsu, H., Kondo, Y., Doi, M., Takeshita, A., Kanno, T. & Maekawa, M. (2003). Detection of mutations in quinolone resistance-determining regions in levofloxacin- and methicillin-resistant *Staphylococcus aureus*: effects of the mutations on fluoroquinolone MICs. *Diagnostic Microbiology and Infectious Disease,* **46,** 139-145.

Horváth, A., Dobay, O., Kardos, S., Ghidán, Á., Tóth, Á., Pászti, J., Ungvári, E., Horváth, P., Nagy, K., Zissman, S. & Füzi, M. (2012). Varying fitness cost associated with resistance to fluoroquinolones governs clonal dynamic of methicillin-resistant *Staphylococcus aureus*. *European Journal of Clinical Microbiology & Infectious Diseases,* **31,** 2029-2036.

Hsu, L. Y., Harris, S. R., Chlebowicz, M. A., Lindsay, J. A., Koh, T. H., Krishnan, P., Tan, T. Y., Hon, P. Y., Grubb, W. B., Bentley, S. D., Parkhill, J., Peacock, S. J. & Holden, M. T. (2015). Evolutionary dynamics of methicillin-resistant *Staphylococcus aureus* within a healthcare system. *Genome Biol.,* **16,** 81.

Hsu, L. Y., Koh, T. H., Singh, K., Kang, M. L., Kurup, A. & Tan, B. H. (2005). Dissemination of multisusceptible methicillin-resistant *Staphylococcus aureus* in Singapore. *J.Clin.Microbiol.,* **43,** 2923-2925.

Hsu, L. Y., Loomba-Chlebicka, N., Koh, T. H., Kang, M. L., Tan, B. H. & Tambyah, P. A. (2007a). EMRSA-15 Bacteremia is not Associated with a Worse Outcome Compared with Bacteremia Caused by Multidrug-Resistant MRSA. *Int.J.Biomed.Sci.,* **3,** 97-103.

Hsu, L. Y., Loomba-Chlebicka, N., Koh, Y. L., Tan, T. Y., Krishnan, P., Lin, R. T., Tee, N. W., Fisher, D. A. & Koh, T. H. (2007b). Evolving EMRSA-15 epidemic in Singapore hospitals. *J.Med.Microbiol.,* **56,** 376-379.

Huang, H., Cohen, S. H., King, J. H., Monchaud, C., Nguyen, H. & Flynn, N. M. (2008). Injecting drug use and community-associated methicillin-resistant *Staphylococcus aureus* infection. *Diagnostic Microbiology and Infectious Disease,* **60,** 347-350.

Huelsenbeck, J. P., Ronquist, F., Nielsen, R. & Bollback, J. P. (2001). Bayesian Inference of Phylogeny and Its Impact on Evolutionary Biology. *Science,* **294,** 2310-2314.

Illumina Incorporate. (2015a). *An Introduction to Next-Generation Sequencing Technology* [Online]. http://www.illumina.com/technology/next-generation-sequencing.html: Illumina Inc. [Accessed 19/02/2016 2016].

Illumina Incorporate. (2015b). *Technology Spotlight: Illumina® Sequencing* [Online]. https://http://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf. [Accessed 19/02/2016 2016].

Ito, T., Ma, X. X., Takeuchi, F., Okuma, K., Yuzawa, H. & Hiramatsu, K. (2004). Novel Type V Staphylococcal Cassette Chromosome mec Driven by a Novel Cassette Chromosome Recombinase, ccrC. *Antimicrobial Agents and Chemotherapy,* **48,** 2637-2651.

Jawad, A., Heritage, J., Snelling, A. M., Gascoyne-Binzi, D. M. & Hawkey, P. M. (1996). Influence of relative humidity and suspending menstrua on survival of Acinetobacter spp. on dry surfaces. *Journal of Clinical Microbiology,* **34,** 2881-2887.

Jeffreys, H. (1935). Some Tests of Significance, Treated by the Theory of Probability. *Mathematical Proceedings of the Cambridge Philosophical Society,* **31,** 203-222.

Jessen, O., Rosendal, K., Bülow, P., Faber, V. & Eriksen, K. R. (1969). Changing Staphylococci and Staphylococcal Infections. *New England Journal of Medicine,* **281,** 627-635.

Jevons, M. P., Coe, A. W. & Parker, M. T. (1963). Methicillin resistance in staphylococci. *Lancet,* **281,** 904-907.

Jiao, W. W., Mokrousov, I., Sun, G. Z., Guo, Y. J., Vyazovaya, A., Narvskaya, O. & Shen, A. D. (2008). Evaluation of New Variable-Number Tandem-Repeat Systems for Typing *Mycobacterium tuberculosis* with Beijing Genotype Isolates from Beijing, China. *Journal of Clinical Microbiology,* **46,** 1045-1049.

Johnson, A. P., Aucken, H. M., Cavendish, S., Ganner, M., Wale, M. C., Warner, M., Livermore, D. M. & Cookson, B. D. (2001). Dominance of EMRSA-15 and -16 among MRSA causing nosocomial bacteraemia in the UK: analysis of isolates from the European Antimicrobial Resistance Surveillance System (EARSS). *J.Antimicrob.Chemother.,* **48,** 143-144.

Kaatz, G. W., Seo, S. M. & Foster, T. J. (1999). Introduction of a norA promoter region mutation into the chromosome of a fluoroquinolone-susceptible strain of *Staphylococcus aureus* using plasmid integration. *Antimicrobial Agents and Chemotherapy,* **43,** 2222-2224.

Kaatz, G. W., Seo, S. M. & Ruble, C. A. (1993). Efflux-mediated fluoroquinolone resistance in *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy,* **37,** 1086-1094.

Kali, A., Stephen, S., Sivaraman, U., Kumar, S., Joseph, N. M., Srirangaraj, S. & Easow, J. M. (2013). Bacteriophage types of methicillin-resistant *Staphylococcus aureus* in a tertiary care hospital. *The Australasian Medical Journal,* **6,** 496-503.

Karanika, S., Kinamon, T., Grigoras, C. & Mylonakis, E. (2016). Colonization With Methicillin-resistant *Staphylococcus aureus* and Risk for Infection Among Asymptomatic Athletes: A Systematic Review and Metaanalysis. *Clinical Infectious Diseases,* **63,** 195-204.

Kass, R. E. & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association,* **90,** 773-795.

Katayama, Y., Ito, T. & Hiramatsu, K. (2000). A New Class of Genetic Element, Staphylococcus Cassette Chromosome mec, Encodes Methicillin Resistance in *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy,* **44,** 1549-1555.

Kato-Maeda, M., Metcalfe, J. Z. & Flores, L. (2011). Genotyping of *Mycobacterium tuberculosis*: application in epidemiologic studies. *Future microbiology,* **6,** 203-216.

Kawamura-Sato, K., Wachino, J.-I., Kondo, T., Ito, H. & Arakawa, Y. (2008). Reduction of disinfectant bactericidal activities in clinically isolated Acinetobacter species in the presence of organic material. *Journal of Antimicrobial Chemotherapy,* **61,** 568-576.

Kayser, F. H. (1975). Methicillin resistant Staphylococci 1965-75. *The Lancet,* **306,** 650-653.

Kennemann, L., Didelot, X., Aebischer, T., Kuhn, S., Drescher, B., Droege, M., Reinhardt, R., Correa, P., Meyer, T. F., Josenhans, C., Falush, D. & Suerbaum, S. (2011). *Helicobacter pylori* genome evolution during human infection. *Proceedings of the National Academy of Sciences,* **108,** 5033-5038.

Kerr, S., Kerr, G. E., Mackintosh, C. A. & Marples, R. R. (1990). A survey of methicillin-resistant *Staphylococcus aureus* affecting patients in England and Wales. *J.Hosp.Infect.,* **16,** 35-48.

Kluytmans, J., Van Belkum, A. & Verbrugh, H. (1997). Nasal carriage of *Staphylococcus aureus:* epidemiology, underlying mechanisms, and associated risks. *Clinical Microbiology Reviews,* **10,** 505-520.

Knight, G. M., Budd, E. L., Whitney, L., Thornley, A., Al-Ghusein, H., Planche, T. & Lindsay, J. A. (2012). Shift in dominant hospital-associated methicillin-resistant *Staphylococcus aureus* (HA-MRSA) clones over time. *J.Antimicrob.Chemother.,* **67,** 2514-2522.

Ko, K. S., Lee, J.-Y., Suh, J. Y., Oh, W. S., Peck, K. R., Lee, N. Y. & Song, J.-H. (2005). Distribution of Major Genotypes among Methicillin-Resistant *Staphylococcus aureus* Clones in Asian Countries. *Journal of Clinical Microbiology,* **43,** 421-426.

Kong, Z., Zhao, P., Liu, H., Yu, X., Qin, Y., Su, Z., Wang, S., Xu, H. & Chen, J. (2016). Whole-Genome Sequencing for the Investigation of a Hospital Outbreak of MRSA in China. *PLoS ONE,* **11,** e0149844.

Koser, C. U., Ellington, M. J., Cartwright, E. J., Gillespie, S. H., Brown, N. M., Farrington, M., Holden, M. T., Dougan, G., Bentley, S. D., Parkhill, J. & Peacock, S. J. (2012a). Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS.Pathog.,* **8,** e1002824.

Koser, C. U., Fraser, L. J., Ioannou, A., Becq, J., Ellington, M. J., Holden, M. T., Reuter, S., Torok, M. E., Bentley, S. D., Parkhill, J., Gormley, N. A., Smith, G. P. & Peacock, S. J. (2014). Rapid single-colony whole-genome sequencing of bacterial pathogens. *J.Antimicrob.Chemother.,* **69,** 1275-1281.

Koser, C. U., Holden, M. T., Ellington, M. J., Cartwright, E. J., Brown, N. M., Ogilvy-Stuart, A. L., Hsu, L. Y., Chewapreecha, C., Croucher, N. J., Harris, S. R., Sanders, M., Enright, M. C., Dougan, G., Bentley, S. D., Parkhill, J., Fraser, L. J., Betley, J. R., Schulz-Trieglaff, O. B.,

Smith, G. P. & Peacock, S. J. (2012b). Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *N.Engl.J.Med.,* **366,** 2267-2275.

Kremer, K., Au, B. K. Y., Yip, P. C. W., Skuce, R., Supply, P., Kam, K. M. & Van Soolingen, D. (2005). Use of Variable-Number Tandem-Repeat Typing To Differentiate Mycobacterium tuberculosis Beijing Family Isolates from Hong Kong and Comparison with IS6110 Restriction Fragment Length Polymorphism Typing and Spoligotyping. *Journal of Clinical Microbiology,* **43,** 314-320.

Kuhn, G., Francioli, P. & Blanc, D. S. (2006). Evidence for Clonal Evolution among Highly Polymorphic Genes in Methicillin-Resistant *Staphylococcus aureus*. *Journal of Bacteriology,* **188,** 169-178.

Lacey, R. W. (1972). Effect of antibiotic resistance on the survival of *Staphylococcus aureus*. *Journal of Clinical Pathology,* **25,** 713-715.

Lacey, R. W., Alder, V. G. & Gillespie, W. A. (1970). The Survival of *Staphylococcus aureus* on Human Skin: An Investigation Using Mixed Cultures. *British Journal of Experimental Pathology,* **51,** 305-313.

Lacey, R. W., Barr, K. W., Barr, V. E. & Inglis, T. J. (1986). Properties of methicillin-resistant *Staphylococcus aureus* colonizing patients in a burns unit. *Journal of Hospital Infection,* **7,** 137-148.

Lafaurie, M., Porcher, R., Donay, J.-L., Touratier, S. & Molina, J.-M. (2012). Reduction of fluoroquinolone use is associated with a decrease in methicillin-resistant *Staphylococcus aureus* and fluoroquinolone-resistant *Pseudomonas aeruginosa* isolation rates: a 10 year study. *Journal of Antimicrobial Chemotherapy,* **67,** 1010-1015.

Landers, T. F., Harris, R. E., Wittum, T. E. & Stevenson, K. B. (2009). Colonization with *Staphylococcus aureus* and Methicillin-Resistant *S. aureus* among a Sample of Homeless Individuals, Ohio. *Infection control and hospital epidemiology : the official journal of the Society of Hospital Epidemiologists of America,* **30,** 801-803.

Lawes, T., Lopez-Lozano, J.-M., Nebot, C. A., Macartney, G., Subbarao-Sharma, R., Dare, C. R. J., Wares, K. D. & Gould, I. M. (2015). Effects of national antibiotic stewardship and infection control strategies on hospital-associated and community-associated meticillin-resistant *Staphylococcus aureus* infections across a region of Scotland: a non-linear time-series study. *The Lancet Infectious Diseases,* **15,** 1438-1449.

Le, S. Q. & Durbin, R. (2011). SNP detection and genotyping from low-coverage sequencing data on multiple diploid samples. *Genome Research,* **21,** 952-960.

Li, H. & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics.,* **26,** 589-595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics.,* **25,** 2078-2079.

Li, Y., Sidore, C., Kang, H. M., Boehnke, M. & Abecasis, G. R. (2011). Low-coverage sequencing: Implications for design of complex trait association studies. *Genome Research,* **21,** 940-951.

Lidwell, O. M. & Lowbury, E. J. (1950). The Survival of Bacteria in Dust: II. The Effect of Atmospheric Humidity on the Survival of Bacteria in Dust. *The Journal of Hygiene,* **48,** 21-27.

Lindsay, J. A. (2013). Hospital-associated MRSA and antibiotic resistance-what have we learned from genomics? *international Journal of Medical Microbiology,* **303,** 318-323.

Lindsay, J. A. (2014). *Staphylococcus aureus* genomics and the impact of horizontal gene transfer. *international Journal of Medical Microbiology,* **304,** 103-109.

Lindsay, J. A., Knight, G. M., Budd, E. L. & Mccarthy, A. J. (2012). Shuffling of mobile genetic elements (MGEs) in successful healthcare-associated MRSA (HA-MRSA). *Mob.Genet.Elements.,* **2,** 239-243.

Liu, G. Y. & Nizet, V. (2009). Color me bad: microbial pigments as virulence factors. *Trends in microbiology,* **17,** 406-413.

Livermore, D. M., James, D., Reacher, M., Graham, C., Nichols, T., Stephens, P., Johnson, A. P. & George, R. C. (2002). Trends in Fluoroquinolone (Ciprofloxacin) Resistance in Enterobacteriaceae from Bacteremias, England and Wales, 1990–1999. *Emerging Infectious Diseases,* **8,** 473-478.

Loman, N. J., Constantinidou, C., Chan, J. Z., Halachev, M., Sergeant, M., Penn, C. W., Robinson, E. R. & Pallen, M. J. (2012). High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat.Rev.Microbiol.,* **10,** 599-606.

Long, S. W., Beres, S. B., Olsen, R. J. & Musser, J. M. (2014). Absence of Patient-to-Patient Intrahospital Transmission of *Staphylococcus aureus* as Determined by Whole-Genome Sequencing. *mBio,* **5,** e01692-14.

Lowy, F. D. (1998). *Staphylococcus aureus* Infections. *New England Journal of Medicine,* **339,** 520-532.

Mackintosh, C. A., Marples, R. R., Kerr, G. E. & Bannister, B. A. (1991). Surveillance of methicillin-resistant *Staphylococcus aureus* in England and Wales, 1986-1990. *J.Hosp.Infect.,* **18,** 279-292.

Marcusson, L. L., Frimodt-Møller, N. & Hughes, D. (2009). Interplay in the Selection of Fluoroquinolone Resistance and Bacterial Fitness. *PLoS Pathog,* **5,** e1000541.

Marples, R. R. & Rosdahl, V. T. (1997). International Quality Control of Phage Typing of *Staphylococcus Aureus*. *Journal of Medical Microbiology,* **46,** 511-516.

Marshall, J. H. & Wilmoth, G. J. (1981). Pigments of *Staphylococcus aureus*, a series of triterpenoid carotenoids. *Journal of Bacteriology,* **147,** 900-913.

Marttinen, P., Hanage, W. P., Croucher, N. J., Connor, T. R., Harris, S. R., Bentley, S. D. & Corander, J. (2011). Detection of recombination events in bacterial genomes from large population samples. *Nucleic Acids Research,* **40,** e6.

Mcadam, P. R., Templeton, K. E., Edwards, G. F., Holden, M. T., Feil, E. J., Aanensen, D. M., Bargawi, H. J., Spratt, B. G., Bentley, S. D., Parkhill, J., Enright, M. C., Holmes, A., Girvan, E. K., Godfrey, P. A., Feldgarden, M., Kearns, A. M., Rambaut, A., Robinson, D. A. & Fitzgerald, J. R. (2012). Molecular tracing of the emergence, adaptation, and transmission of hospital-associated methicillin-resistant *Staphylococcus aureus*. *Proc.Natl.Acad.Sci.U.S.A,* **109,** 9107-9112.

Mcdade, J. J. & Hall, L. B. (1964). Survival of *Staphylococcus aureus* in the environment:II. Effect of elevated temperature on surface exposed staphylococci. *American Journal of Epidemiology,* **80,** 184-191.

Mcdonald, M., Dougall, A., Holt, D., Huygens, F., Oppedisano, F., Giffard, P. M., Inman-Bamber, J., Stephens, A. J., Towers, R., Carapetis, J. R. & Currie, B. J. (2006). Use of a Single-Nucleotide Polymorphism Genotyping System To Demonstrate the Unique Epidemiology of Methicillin-Resistant *Staphylococcus aureus* in Remote Aboriginal Communities. *Journal of Clinical Microbiology,* **44,** 3720-3727.

Mcginley, K. J., Larson, E. L. & Leyden, J. J. (1988). Composition and density of microflora in the subungual space of the hand. *Journal of Clinical Microbiology,* **26,** 950-953.

Mertz, D., Frei, R., Periat, N. & Et Al. (2009). Exclusive *Staphylococcus aureus* throat carriage: At-risk populations. *Archives of Internal Medicine,* **169,** 172-178.

Metzker, M. L. (2005). Emerging technologies in DNA sequencing. *Genome Research,* **15,** 1767-1776.

Miller, R. R., Walker, A. S., Godwin, H., Fung, R., Votintseva, A., Bowden, R., Mant, D., Peto, T. E., Crook, D. W. & Knox, K. (2014). Dynamics of acquisition and loss of carriage of *Staphylococcus aureus* strains in the community: The effect of clonal complex. *Journal of Infection,* **68,** 426-439.

Murchan, S., Kaufmann, M. E., Deplano, A., De Ryck, R., Struelens, M., Zinn, C. E., Fussing, V., Salmenlinna, S., Vuopio-Varkila, J., El Solh, N., Cuny, C., Witte, W., Tassios, P. T., Legakis, N., Van Leeuwen, W., Van Belkum, A., Vindel, A., Laconcha, I., Garaizar, J., Haeggman, S., Olsson-Liljequist, B., Ransjo, U., Coombes, G. & Cookson, B. (2003). Harmonization of Pulsed-Field Gel Electrophoresis Protocols for Epidemiological Typing of Strains of Methicillin-Resistant *Staphylococcus aureus*: a Single Approach Developed by Consensus in 10 European Laboratories and Its Application for Tracing the Spread of Related Strains. *Journal of Clinical Microbiology,* **41,** 1574-1585.

Nathwani, D., Sneddon, J., Malcolm, W., Wiuff, C., Patton, A., Hurding, S., Eastaway, A., Seaton, R. A., Watson, E., Gillies, E., Davey, P. & Bennie, M. (2011). Scottish Antimicrobial Prescribing Group (SAPG): development and impact of the Scottish National Antimicrobial Stewardship Programme. *International Journal of Antimicrobial Agents,* **38,** 16-26.

Neely, A. N. & Maley, M. P. (2000). Survival of Enterococci and Staphylococci on Hospital Fabrics and Plastic. *Journal of Clinical Microbiology,* **38,** 724-726.

Noble, W. C. (1962). The dispersal of staphylococci in hospital wards. *Journal of Clinical Pathology,* **15,** 552-558.

Nocker, A., Fernández, P. S., Montijn, R. & Schuren, F. (2012). Effect of air drying on bacterial viability: A multiparameter viability assessment. *Journal of Microbiological Methods,* **90,** 86-95.

Nubel, U., Nachtnebel, M., Falkenhorst, G., Benzler, J., Hecht, J., Kube, M., Brocker, F., Moelling, K., Buhrer, C., Gastmeier, P., Piening, B., Behnke, M., Dehnert, M., Layer, F., Witte, W. & Eckmanns, T. (2013). MRSA transmission on a neonatal intensive care unit: epidemiological and genome-based phylogenetic analyses. *PLoS.One.,* **8,** e54898.

Ochman, H. & Davalos, L. M. (2006). The nature and dynamics of bacterial genomes. *Science,* **311,** 1730-1733.

Okuma, K., Iwakawa, K., Turnidge, J. D., Grubb, W. B., Bell, J. M., O'brien, F. G., Coombs, G. W., Pearman, J. W., Tenover, F. C., Kapi, M., Tiensasitorn, C., Ito, T. & Hiramatsu, K. (2002). Dissemination of New Methicillin-Resistant *Staphylococcus aureus* Clones in the Community. *Journal of Clinical Microbiology,* **40,** 4289-4294.

Otter, J. A., S., Y. & French, G. L. (2011). The role played by contaminated surfaces in the transmission of nosocomial pathogens. *Infect Control Hosp Epidemiology,* **32,** 687-699.

Otto, M. (2014). *Staphylococcus aureus* toxins. *Current Opinion in Microbiology,* **17,** 32-37.

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., Fookes, M., Falush, D., Keane, J. A. & Parkhill, J. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics,* **31,** 3691-3693.

Pan, E. S., Diep, B. A., Carleton, H. A., Charlebois, E. D., Sensabaugh, G. F., Haller, B. L. & Remington, F. P. (2003). Increasing Prevalence of Methicillin-Resistant *Staphylococcus aureus* Infection in California Jails. *Clinical Infectious Diseases,* **37,** 1384-1388.

Parienti, J. J., Cattoir, V., Thibon, P., Lebouvier, G., Verdon, R., Daubin, C., Du Cheyron, D., Leclercq, R. & Charbonneau, P. (2011). Hospital-wide modification of fluoroquinolone policy and meticillin-resistant *Staphylococcus aureus* rates: a 10-year interrupted time-series analysis. *Journal of Hospital Infection,* **78,** 118-122.

Patel, R. K. & Jain, M. (2012). NGS QC Toolkit: A Toolkit for Quality Control of Next Generation Sequencing Data. *PLOS ONE,* **7,** e30619.

Paterson, G. K., Harrison, E. M., Murray, G. G., Welch, J. J., Warland, J. H., Holden, M. T., Morgan, F. J., Ba, X., Koop, G., Harris, S. R., Maskell, D. J., Peacock, S. J., Herrtage, M. E., Parkhill, J. & Holmes, M. A. (2015). Capturing the cloud of diversity reveals complexity and heterogeneity of MRSA carriage, infection and transmission. *Nat.Commun.,* **6,** 6560.

Pearson, T., Okinaka, R. T., Foster, J. T. & Keim, P. (2009). Phylogenetic understanding of clonal populations in an era of whole genome sequencing. *Infection, Genetics and Evolution,* **9,** 1010-1019.

Pelz, A., Wieland, K.-P., Putzbach, K., Hentschel, P., Albert, K. & Götz, F. (2005). Structure and Biosynthesis of Staphyloxanthin from *Staphylococcus aureus*. *Journal of Biological Chemistry,* **280,** 32493-32498.

Peter, B. J., Arsuaga, J., Breier, A. M., Khodursky, A. B., Brown, P. O. & Cozzarelli, N. R. (2004). Genomic transcriptional response to loss of chromosomal supercoiling in *Escherichia coli*. *Genome Biology,* **5,** R87-R87.

Pettit, F. & Lowbury, E. J. (1968). Survival of wound pathogens under different environmental conditions. *The Journal of Hygiene,* **66,** 393-406.

Piddock, L. J. V. (1999). Mechanisms of Fluoroquinolone Resistance: An Update 1994–1998. *Drugs,* **58,** 11-18.

Planet, P. J., Narechania, A., Chen, L., Mathema, B., Boundy, S., Archer, G. & Kreiswirth, B. (2016). Architecture of a Species: Phylogenomics of *Staphylococcus aureus*. *Trends in Microbiology***,** 153-166.

Posada, D. & Crandall, K. A. (2002). The Effect of Recombination on the Accuracy of Phylogeny Estimation. *Journal of Molecular Evolution,* **54,** 396-402.

Potts, M. (1994). Desiccation tolerance of prokaryotes. *Microbiological Reviews,* **58,** 755-805.

Price, J. R., Didelot, X., Crook, D. W., Llewelyn, M. J. & Paul, J. (2013). Whole genome sequencing in the prevention and control of *Staphylococcus aureus* infection. *J.Hosp.Infect.,* **83,** 14-21.

Price, J. R., Golubchik, T., Cole, K., Wilson, D. J., Crook, D. W., Thwaites, G. E., Bowden, R., Walker, A. S., Peto, T. E., Paul, J. & Llewelyn, M. J. (2014). Whole-Genome Sequencing Shows That Patient-to-Patient Transmission Rarely Accounts for Acquisition of *Staphylococcus aureus* in an Intensive Care Unit. *Clinical Infectious Diseases***,** 609-618.

Prosperi, M., Veras, N., Azarian, T., Rathore, M., Nolan, D., Rand, K., Cook, R. L., Johnson, J., Morris, J. G. & Salemi, M. (2013). Molecular Epidemiology of Community-Associated Methicillin-resistant Staphylococcus aureus in the genomic era: a Cross-Sectional Study. *Scientific Reports,* **3,** 1902.

Public Health England (2015). English surveillance programme for antimicrobial utilisation and resistance (ESPAUR) 2010 to 2014. http://www.gov.uk/phe.

Public Health England (2016). MRSA bacteraemia: Annual Data. http://www.gov.uk/phe.

Reacher, M. H., Shah, A., Livermore, D. M., Wale, M. C. J., Graham, C., Johnson, A. P., Heine, H., Monnickendam, M. A., Barker, K. F., James, D. & George, R. C. (2000). Bacteraemia and

antibiotic resistance of its pathogens reported in England and Wales between 1990 and 1998: trend analysis. *BMJ : British Medical Journal,* **320,** 213-216.

Redgrave, L. S., Sutton, S. B., Webber, M. A. & Piddock, L. J. (2014). Fluoroquinolone resistance: mechanisms, impact on bacteria, and role in evolutionary success. *Trends Microbiol.,* **22,** 438-445.

Reed, K., Stemper, M. & Shukla, S. 2007. Pulsed-Field Gel Electrophoresis of MRSA. *In:* JI, Y. (ed.) *Methicillin-Resistant Staphylococcus aureus (MRSA) Protocols.* Humana Press.

Reuter, S., Török, E. M., Holden, M. T., Reynolds, R., Raven, K. E., Blane, B., Donker, T., Bentley, S. D., Aanensen, D. M., Grundmann, H., Feil, E. J., Spratt, B. G., Parkhill, J. & Peacock, S. J. (2015). Building a genomic framework for prospective MRSA surveillance in the United Kingdom and the Republic of Ireland. *Genome Research*.

Reuter, S., Török, M. E., Holden, M. T. G., Reynolds, R., Raven, K. E., Blane, B., Donker, T., Bentley, S. D., Aanensen, D. M., Grundmann, H., Feil, E. J., Spratt, B. G., Parkhill, J. & Peacock, S. J. (2016). Building a genomic framework for prospective MRSA surveillance in the United Kingdom and the Republic of Ireland. *Genome Research,* **26,** 263-270.

Richardson, E. J. & Watson, M. (2013). The automatic annotation of bacterial genomes. *Brief.Bioinform.,* **14,** 1-12.

Richardson, J. F. & Reith, S. (1993). Characterization of a strain of methicillin-resistant *Staphylococcus aureus* (EMRSA-15) by conventional and molecular methods. *J.Hosp.Infect.,* **25,** 45-52.

Rosenbach, F. J. (1884). Mikroorganismenbeiden Wundinfektionskrankheiten des Menschen. *J. F. Berg- man,Wiesbaden.*

Rosendal, K., Bülow, P., Bentzon, M. W. & Eriksen, K. R. (1976). *Staphylococcus aureus* strains isolated in Danish hospitals from 1st, 1966, to December 31st, 1974. *Acta Pathologica Microbiologica Scandinavica Section B Microbiology,* **84B,** 359-368.

Rountree, P. M. (1963). The effect of desiccation on the viability of *Staphylococcus aureus*. *The Journal of Hygiene,* **61,** 265-272.

Rozen, D. E., Mcgee, L., Levin, B. R. & Klugman, K. P. (2007). Fitness Costs of Fluoroquinolone Resistance in *Streptococcus pneumoniae*. *Antimicrobial Agents and Chemotherapy,* **51,** 412-416.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A. & Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics.,* **16,** 944-945.

Salipante, S. J., Sengupta, D. J., Cummings, L. A., Land, T. A., Hoogestraat, D. R. & Cookson, B. T. (2015). Application of Whole-Genome Sequencing for Bacterial Strain Typing in Molecular Epidemiology. *Journal of Clinical Microbiology,* **53,** 1072-1079.

Sarma, J. B., Marshall, B., Cleeve, V., Tate, D., Oswald, T. & Woolfrey, S. (2015). Effects of fluoroquinolone restriction (from 2007 to 2012) on *Clostridium difficile* infections: interrupted time-series analysis. *Journal of Hospital Infection,* **91,** 74-80.

Schmid-Hempel, P. & Frank, S. A. (2007). Pathogenesis, Virulence, and Infective Dose. *PLoS Pathog,* **3,** e147.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics,* **30,** 2068-2069.

Shang, W., Hu, Q., Yuan, W., Cheng, H., Yang, J., Hu, Z., Yuan, J., Zhang, X., Peng, H., Yang, Y., Hu, X., Li, M., Zhu, J. & Rao, X. (2015). Comparative Fitness and Determinants for the Characteristic Drug Resistance of ST239-MRSA-III-t030 and ST239-MRSA-III-t037 Strains Isolated in China. *Microbial Drug Resistance,* **22**.

Shaw, L. N., Lindholm, C., Prajsnar, T. K., Miller, H. K., Brown, M. C., Golonka, E., Stewart, G. C., Tarkowski, A. & Potempa, J. (2008). Identification and Characterization of σ(S), a Novel Component of the *Staphylococcus aureus* Stress and Virulence Responses. *PLoS ONE,* **3,** e3844.

Shopsin, B., Gomez, M., Montgomery, S. O., Smith, D. H., Waddington, M., Dodge, D. E., Bost, D. A., Riehman, M., Naidich, S. & Kreiswirth, B. N. (1999). Evaluation of Protein A Gene Polymorphic Region DNA Sequencing for Typing of *Staphylococcus aureus* Strains. *Journal of Clinical Microbiology,* **37,** 3556-3563.

Silverman, S. H., Johnson, M., Burdon, D. W. & Keighley, M. R. B. (1986). Pharmacokinetics of single dose intravenous ciprofloxacin in patients undergoing gastrointestinal surgery. *Journal of Antimicrobial Chemotherapy,* **18,** 107-112.

Simões, M., Pereira, M. O., Machado, I., Simões, L. C. & Vieira, M. J. (2006). Comparative antibacterial potential of selected aldehyde-based biocides and surfactants against planktonic Pseudomonas fluorescens. *Journal of Industrial Microbiology and Biotechnology,* **33,** 741-749.

Simon, H. J. (1962). The Newer Penicillins. *California Medicine,* **97,** 135-141.

Sims, D., Sudbery, I., Ilott, N. E., Heger, A. & Ponting, C. P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. *Nat.Rev.Genet.,* **15,** 121-132.

Skinner, D. & Keefer, C. S. (1941). Significance of bacteremia caused by *Staphylococcus aureus*: A study of one hundred and twenty-two cases and a review of the literature concerned with experimental infection in animals. *Archives of Internal Medicine,* **68,** 851-875.

Song, J.-H., Hsueh, P.-R., Chung, D. R., Ko, K. S., Kang, C.-I., Peck, K. R., Yeom, J.-S., Kim, S.-W., Chang, H.-H., Kim, Y.-S., Jung, S.-I., Son, J. S., So, T. M.-K., Lalitha, M. K., Yang, Y., Huang, S.-G., Wang, H., Lu, Q., Carlos, C. C., Perera, J. A., Chiu, C.-H., Liu, J.-W., Chongthaleong, A., Thamlikitkul, V. & Van, P. H. (2011). Spread of methicillin-resistant *Staphylococcus*

aureus between the community and the hospitals in Asian countries: an ANSORP study. *Journal of Antimicrobial Chemotherapy,* **66,** 1061-1069.

Souverein, D., Houtman, P., Euser, S. M., Herpers, B. L., Kluytmans, J. & Den Boer, J. W. (2016). Costs and Benefits Associated with the MRSA Search and Destroy Policy in a Hospital in the Region Kennemerland, The Netherlands. *PLOS ONE,* **11,** e0148175.

Stefani, S., Chung, D. R., Lindsay, J. A., Friedrich, A. W., Kearns, A. M., Westh, H. & Mackenzie, F. M. (2012). Meticillin-resistant *Staphylococcus aureus* (MRSA): global epidemiology and harmonisation of typing methods. *International Journal of Antimicrobial Agents,* **39,** 273-282.

Supply, P., Allix, C., Lesjean, S., Cardoso-Oelemann, M., Rüsch-Gerdes, S., Willery, E., Savine, E., De Haas, P., Van Deutekom, H., Roring, S., Bifani, P., Kurepina, N., Kreiswirth, B., Sola, C., Rastogi, N., Vatin, V., Gutierrez, M. C., Fauville, M., Niemann, S., Skuce, R., Kremer, K., Locht, C. & Van Soolingen, D. (2006). Proposal for Standardization of Optimized Mycobacterial Interspersed Repetitive Unit-Variable-Number Tandem Repeat Typing of Mycobacterium tuberculosis. *Journal of Clinical Microbiology,* **44,** 4498-4510.

Supply, P., Mazars, E., Lesjean, S., Vincent, V., Gicquel, B. & Locht, C. (2000). Variable human minisatellite-like regions in the Mycobacterium tuberculosis genome. *Molecular Microbiology,* **36,** 762-771.

Tacconelli, E., De Angelis, G., Cataldo, M. A., Pozzi, E. & Cauda, R. (2008). Does antibiotic exposure increase the risk of methicillin-resistant *Staphylococcus aureus* (MRSA) isolation? A systematic review and meta-analysis. *Journal of Antimicrobial Chemotherapy,* **61,** 26-38.

Tajeddin, E., Rashidan, M., Razaghi, M., Javadi, S. S. S., Sherafat, S. J., Alebouyeh, M., Sarbazi, M. R., Mansouri, N. & Zali, M. R. (2016). The role of the intensive care unit environment and health-care workers in the transmission of bacteria associated with hospital acquired infections. *Journal of Infection and Public Health,* **9,** 13-23.

Takahashi, H., Kikuchi, T., Shoji, S., Fujimura, S., Lutfor, A. B., Tokue, Y., Nukiwa, T. & Watanabe, A. (1998). Characterization of *gyrA, gyrB, grlA* and *grlB* mutations in fluoroquinolone-resistant clinical isolates of *Staphylococcus aureus*. *Journal of Antimicrobial Chemotherapy,* **41,** 49-57.

Teare, L., Shelley, O. P., Millership, S. & Kearns, A. (2010). Outbreak of Panton–Valentine leucocidin-positive meticillin-resistant *Staphylococcus aureus* in a regional burns unit. *Journal of Hospital Infection,* **76,** 220-224.

Tenover, F. C., Arbeit, R. D., Goering, R. V., Mickelsen, P. A., Murray, B. E., Persing, D. H. & Swaminathan, B. (1995). Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *Journal of Clinical Microbiology,* **33,** 2233-2239.

Thouverez, M., Muller, A., Hocquet, D., Talon, D. & Bertrand, X. (2003). Relationship between molecular epidemiology and antibiotic susceptibility of methicillin-resistant

Staphylococcus aureus (MRSA) in a French teaching hospital. *Journal of Medical Microbiology,* **52,** 801-806.

Tong, S. Y., Holden, M. T., Nickerson, E. K., Cooper, B. S., Koser, C. U., Cori, A., Jombart, T., Cauchemez, S., Fraser, C., Wuthiekanun, V., Thaipadungpanit, J., Hongsuwan, M., Day, N. P., Limmathurotsakul, D., Parkhill, J. & Peacock, S. J. (2015a). Genome sequencing defines phylogeny and spread of methicillin-resistant *Staphylococcus aureus* in a high transmission setting. *Genome Res.,* **25,** 111-118.

Tong, S. Y. C., Davis, J. S., Eichenberger, E., Holland, T. L. & Fowler, V. G. (2015b). Staphylococcus aureus Infections: Epidemiology, Pathophysiology, Clinical Manifestations, and Management. *Clinical Microbiology Reviews,* **28,** 603-661.

Tuladhar, E., Hazeleger, W. C., Koopmans, M., Zwietering, M. H., Beumer, R. R. & Duizer, E. (2012). Residual Viral and Bacterial Contamination of Surfaces after Cleaning and Disinfection. *Applied and Environmental Microbiology,* **78,** 7769-7775.

Udo, E. E., Al-Lawati, B. A., Al-Muharmi, Z. & Thukral, S. S. (2014). Genotyping of methicillin-resistant *Staphylococcus aureus* in the Sultan Qaboos University Hospital, Oman reveals the dominance of Panton-Valentine leucocidin-negative ST6-IV/t304 clone. *New Microbes.New Infect.,* **2,** 100-105.

Udo, E. E., Al-Sweih, N. & Noronha, B. (2006). Characterisation of non-multiresistant methicillin-resistant Staphylococcus aureus (including EMRSA-15) in Kuwait Hospitals. *Clin.Microbiol.Infect.,* **12,** 262-269.

Udo, E. E., Boswihi, S. S. & Al-Sweih, N. (2016). High prevalence of toxic shock syndrome toxin–producing epidemic methicillin-resistant Staphylococcus aureus 15 (EMRSA-15) strains in Kuwait hospitals. *New Microbes and New Infections,* **12,** 24-30.

Udo, E. E., Pearman Jw Fau - Grubb, W. B. & Grubb, W. B. Genetic analysis of community isolates of methicillin-resistant Staphylococcus aureus in Western Australia.

Uhlemann, A.-C., Dordel, J., Knox, J. R., Raven, K. E., Parkhill, J., Holden, M. T. G., Peacock, S. J. & Lowy, F. D. (2014). 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,* **111,** 6738-6743.

Une, T., Fujimoto, T., Sato, K. & Osada, Y. (1988). In vitro activity of DR-3355, an optically active ofloxacin. *Antimicrobial Agents and Chemotherapy,* **32,** 1336-1340.

Van Belkum, A., Verkaik, N. J., De Vogel, C. P., Boelens, H. A., Verveer, J., Nouwen, J. L., Verbrugh, H. A. & Wertheim, H. F. (2009). Reclassification of *Staphylococcus aureus* nasal carriage types. *J.Infect.Dis.,* **199,** 1820-1826.

Van Rijen, M. M. L. & Kluytmans, J. a. J. W. (2014). Adjustment of the MRSA Search and Destroy policy for outpatients in the Netherlands: a prospective cohort study with repeated prevalence measurements. *Antimicrobial Resistance and Infection Control,* **3,** 3-3.

Vandenesch, F., Naimi, T., Enright, M. C., Lina, G., Nimmo, G. R., Heffernan, H., Liassine, N., Bes, M., Greenland, T., Reverdy, M.-E. & Etienne, J. (2003). Community-Acquired Methicillin-Resistant *Staphylococcus aureus* Carrying Panton-Valentine Leukocidin Genes: Worldwide Emergence. *Emerging Infectious Diseases,* **9,** 978-984.

Wagenvoort, J. H. T., Sluijsmans, W. & Penders, R. J. R. (2000). Better environmental survival of outbreak vs. sporadic MRSA isolates. *Journal of Hospital Infection,* **45,** 231-234.

Walker, T. M., Ip, C. L., Harrell, R. H., Evans, J. T., Kapatai, G., Dedicoat, M. J., Eyre, D. W., Wilson, D. J., Hawkey, P. M., Crook, D. W., Parkhill, J., Harris, D., Walker, A. S., Bowden, R., Monk, P., Smith, E. G. & Peto, T. E. (2013). Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect.Dis.,* **13,** 137-146.

Wang, J., Wang, W., Li, R., Li, Y., Tian, G., Goodman, L., Fan, W., Zhang, J., Li, J., Zhang, J., Guo, Y., Feng, B., Li, H., Lu, Y., Fang, X., Liang, H., Du, Z., Li, D., Zhao, Y., Hu, Y., Yang, Z., Zheng, H., Hellmann, I., Inouye, M., Pool, J., Yi, X., Zhao, J., Duan, J., Zhou, Y., Qin, J., Ma, L., Li, G., Yang, Z., Zhang, G., Yang, B., Yu, C., Liang, F., Li, W., Li, S., Li, D., Ni, P., Ruan, J., Li, Q., Zhu, H., Liu, D., Lu, Z., Li, N., Guo, G., Zhang, J., Ye, J., Fang, L., Hao, Q., Chen, Q., Liang, Y., Su, Y., San, A., Ping, C., Yang, S., Chen, F., Li, L., Zhou, K., Zheng, H., Ren, Y., Yang, L., Gao, Y., Yang, G., Li, Z., Feng, X., Kristiansen, K., Wong, G. K.-S., Nielsen, R., Durbin, R., Bolund, L., Zhang, X., Li, S., Yang, H. & Wang, J. (2008). The diploid genome sequence of an Asian individual. *Nature,* **456,** 60-65.

Ward, M. J., Gibbons, C. L., Mcadam, P. R., Van Bunnik, B. a. D., Girvan, E. K., Edwards, G. F., Fitzgerald, J. R. & Woolhouse, M. E. J. (2014). Time-Scaled Evolutionary Analysis of the Transmission and Antibiotic Resistance Dynamics of *Staphylococcus aureus* Clonal Complex 398. *Applied and Environmental Microbiology,* **80,** 7275-7282.

Webber, M. A., Ricci, V., Whitehead, R., Patel, M., Fookes, M., Ivens, A. & Piddock, L. J. V. (2013). Clinically Relevant Mutant DNA Gyrase Alters Supercoiling, Changes the Transcriptome, and Confers Multidrug Resistance. *mBio,* **4,** e00273-13.

Weber, S. G., Gold, H. S., Hooper, D. C., Karchmer, A. W. & Carmeli, Y. (2003). Fluoroquinolones and the Risk for Methicillin-resistant *Staphylococcus aureus* in Hospitalized Patients. *Emerging Infectious Diseases,* **9,** 1415-1422.

Wentworth, B. B. (1963). Bacteriophage typing of the Staphylococci. *Bacteriological Reviews,* **27,** 253-272.

Williams, R. E. O. (1963). Healthy carriage of Staphylococcus aureus:it's prevalence and importance. *Bacteriological Reviews,* **27,** 56-71.

Williams, R. E. O. & Dean, B. A. (1974). Phage types of Staphylococcus aureus in one hospital 1961-72. *The Journal of Hygiene,* **73,** 375-382.

Wiśniewska, K., Szewczyk, A., Piechowicz, L., Bronk, M., Samet, A. & Świeć, K. (2012). The use of *spa* and phage typing for characterization of clinical isolates of methicillin-

resistant *Staphylococcus aureus* in the University Clinical Center in Gdańsk, Poland. *Folia Microbiologica,* **57,** 243-249.

Witte, W., Enright, M., Schmitz, F. J., Cuny, C., Braulke, C. & Heuck, D. (2001). Characteristics of a new epidemic MRSA in Germany ancestral to United Kingdom EMRSA 15. *Int.J.Med.Microbiol.,* **290,** 677-682.

World Healthcare Organisation. (2009). WHO guidelines on hand hygiene in health care.

Yang, Z. & Rannala, B. (2012). Molecular phylogenetics: principles and practice. *Nat Rev Genet,* **13,** 303-314.

Yoshida, H., Bogaki, M., Nakamura, S., Ubukata, K. & Konno, M. (1990). Nucleotide sequence and characterization of the *Staphylococcus aureus norA* gene, which confers resistance to quinolones. *Journal of Bacteriology,* **172,** 6942-6949.

Young, B. C., Golubchik, T., Batty, E. M., Fung, R., Larner-Svensson, H., Votintseva, A. A., Miller, R. R., Godwin, H., Knox, K., Everitt, R. G., Iqbal, Z., Rimmer, A. J., Cule, M., Ip, C. L., Didelot, X., Harding, R. M., Donnelly, P., Peto, T. E., Crook, D. W., Bowden, R. & Wilson, D. J. (2012). Evolutionary dynamics of *Staphylococcus aureus* during progression from carriage to disease. *Proc.Natl.Acad.Sci.U.S.A,* **109,** 4550-4555.

Zinderman, C. E., Conner, B., Malakooti, M. A., Lamar, J. E., Armstrong, A. & Bohnker, B. K. (2004). Community-acquired methicillin-resistant *Staphylococcus aureus* among military recruits. *Emerg.Infect.Dis.,* **10,** 941-944.