THE UNIVERSITY
of EDINBURGH

Population analysis of bacterial pathogens on distinct temporal and spatial scales

Paul McAdam

Thesis presented for the degree of Doctor of Philosophy

The University of Edinburgh

2013

The research presented in this thesis is entirely my own work, except where otherwise stated. No part of this thesis has been submitted in any other application for a degree or professional qualification.

…….

Paul McAdam August 2013

Acknowledgements

I would like to thank Ross Fitzgerald for his guidance and support as a supervisor throughout my PhD. I would also like to thank members of the Laboratory for Bacterial Evolution and Pathogenesis, and members of Epigroup for their input and advice.

In addition, I am very grateful to Kate Templeton, Giles Edwards, Ashley Robinson, Andrew Rambaut, and Matt Holden for their contributions to this work, and for many fruitful discussions on pathogen epidemiology and evolution.

Isolates for these studies were kindly provided by C. Branger, P. Schlievert, M. Enright, A. Kearns, R. Goering, H. de Lencastre, A. Shore, E. Feil, D. Aanensen, D. Lindsay, and T. Harrison. Sequencing services were provided by ARK-Genomics (The Roslin Institute).

Finally, I have to show my appreciation to my friends and family, for reminding me there's more to life than research. Thanks to Laura, mam, Jim, Ross, Lyndon, Mhairi, Ben, Owen, and Renee for their encouragement and support over the past 4 years.

Abbreviations

AIDS	Acquired immune deficiency syndrome
BCYE	Buffered charcoal yeast extract
BEAST	Bayesian evolutionary analysis by sampling trees
BLAST	Basic local alignment search tool
bp	Base pair
BSSVS	Bayesian stochastic search variable selection
CA	Community associated/acquired
CC	Clonal complex
CDM	Chemically defined media
CDS	Coding DNA sequence
CF	Cystic fibrosis
CFTR	Cystic fibrosis transmembrane conductance regulator
CFU	Colony forming unit
DNA	Deoxyribonucleic acid
ELISA	Enzyme-linked immunosorbent assay
EMRSA	Epidemic methicillin resistant *Staphylococcus aureus*
ERIC	Enterobacterial repetitive intergenic consensus
GARD	Genetic algorithm for recombination detection
GOLD	Genomes online database
GTR	General time reversible
HA	Hospital associated/acquired
HIV	Human immunodeficiency virus
HKY	Hasegawa Kishino Yano
HPD	Highest posterior density
HUS	Haemolytic uraemic syndrome
kb	Kilobase
LD	Legionnaires' disease
mAb	Monoclonal antibody
Mb	Megabase
MCMC	Markov chain Monte Carlo
MEP	Measurably evolving population
MGE	Mobile genetic element

MIC	Minimum inhibitory concentration
MLST	Multi-locus sequence typing
MRCA	Most recent common ancestor
MRSA	Methicillin resistant *Staphylococcus aureus*
MSSA	Methicillin sensitive *Staphylococcus aureus*
NCBI	National Center for Biotechnology Information
OD	Optical density
PAUP	Phylogenetic analysis using parsimony
PBP2a	Penicillin binding protein 2a
PCR	Polymerase chain reaction
PFGE	Pulsed field gel electrophoresis
PHE	Public Health England
PVL	Panton-Valentine leukocidin
RAxML	Randomized Accelerated Maximum Likelihood
RNA	Ribonucleic acid
rpm	Revolutions per minute
SaPI	Staphylococcal pathogenicity island
SBT	Sequence based typing
SCC	Staphylococcal cassette chromosome
SCV	Small colony variant
Sg	Serogroup
SRA	Sequence read archive
ST	Sequence type
SWP	Southwest Pacific
T4SS	Type 4 secretion system
TSA	Tryptic soy agar
TSB	Tryptic soy broth
TSS	Toxic shock syndrome
TSST-1	Toxic shock syndrome toxin 1
WGS	Whole genome sequencing

Contents

McAdam, PR, Holmes, A, Templeton, KE & Fitzgerald, JR. **(**2011**).** Adaptive Evolution of *Staphylococcus aureus* during Chronic Endobronchial Infection of a Cystic Fibrosis Patient. *PLoS One* 6(9):e24301.

McAdam, PR, Templeton, KE, Edwards, GF, Holden, MTG, Feil, EJ, Aanensen, DM, Bargawi, HJA, Spratt, BG, Bentley, SD, Parkhill, J, Enright, MC, Holmes, A, Girvan, EK, Godfrey, PA, Feldgarden, M, Kearns, AM, Rambaut, A, Robinson, DA & Fitzgerald, JR. **(**2012**).** Molecular tracing of the emergence, adaptation, and transmission of hospital-associated methicillin-resistant *Staphylococcus aureus*. *Proc Natl Acad Sci* 109(23): 9107–12.

Holden, MTG, Hsu, L-Y, Kurt, K, Weinert, LA, Mather, AE, Harris, SR, Strommenger, B, Layer, F, Witte, W, de Lencastre, H, Skov, R, Westh, H, Zemlickova, H, Coombs, G, Kearns, AM, Hill, RLR, Edgeworth, J, Gould, I, Gant, V, Cooke, J, Edwards, GF, McAdam, PR, Templeton, KE, McCann, A, Zhou, Z, Castillo-Ramirez, S, Feil, EJ, Hudson, LO, Enright, MC, Balloux, F, Aanensen, DM, Spratt, BG, Fitzgerald, JR, Parkhill, J, Achtman, M, Bentley, SD, Nübel, U, Zemlicková, H & Castillo-Ramírez, S. **(**2013**).** A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant *Staphylococcus aureus* pandemic. *Genome Res* 23(4):653–664.

Kreutzfeldt, KM, McAdam, PR, Claxton, P, Holmes, A, Seagar, AL, Laurenson, IF & Fitzgerald, JR. **(**2013**).** Molecular Longitudinal Tracking of *Mycobacterium abscessus* spp. during Chronic Infection of the Human Lung. *PLoS One* 16(8):e63237.

Holmes, A, McAllister, G, McAdam, PR, Hsien Choi, S, Girvan, K, Robb, A, Edwards, G, Templeton, K & Fitzgerald, JR. **(**2014**).** Genome-wide single

nucleotide polymorphism-based assay for high-resolution epidemiological analysis of the methicillin-resistant *Staphylococcus aureus* hospital clone EMRSA-15. *Clin Microbiol Infect* 20(2):O124–O131

Spoor, LE, McAdam, PR, Weinert, LA, Rambaut, A, Hasman, H, Aarestrup, FM, Kearns, AM, Larsen, AR, Skov, RL & Fitzgerald, JR. **(**2013**).** Livestock Origin for a Human Pandemic Clone of Community-Associated Methicillin-Resistant *Staphylococcus aureus*. *MBio* 4(4):e00356–13.

LIST OF TABLES

Chapter 1

Chapter 2

Chapter 4

LIST OF FIGURES

Chapter 2

Chapter 3

Chapter 4

Abstract

Bacteria have been the causative agents of major infectious disease pandemics throughout human history. Over the past 4 decades, a combination of changing medical practices, industrialization, and globalisation have led to a number of emergences and re-emergences of bacterial pathogens. The design of rational control programs and bespoke therapies will require an enhanced understanding of the dynamics underpinning the emergence and transmission of pathogenic clones. The recent development of new technologies for sequencing bacterial genomes rapidly and economically has led to a greatly enhanced understanding of the diversity of bacterial populations.

This thesis describes the application of whole genome sequencing of 2 bacterial pathogens, *Staphylococcus aureus* and *Legionella pneumophila*, in order to understand the dynamics of bacterial infections on different temporal and spatial scales. The first study involves the examination of *S. aureus* evolution during a chronic infection of a single patient over a period of 26 months revealing differences in antibiotic resistance profiles and virulence factor expression over time. The genetic variation identified correlated with differences in growth rate, haemolytic activity, and antibiotic sensitivity, implying a profound effect on the ecology of *S. aureus*. Importantly, polymorphisms were identified in global regulators of virulence, with a high frequency of polymorphisms within the SigB locus identified, suggesting this region may be under selection in this patient. The identification of genes under diversifying selection during long-term infection may inform the design of novel therapeutics for the control of refractory chronic infections.

Secondly, the emergence and transmission of 3 pandemic lineages derived from *S. aureus* clonal complex 30 (CC30) were investigated. Independent origins for each pandemic lineage were identified, with striking molecular correlates of hospital- or community-associated pandemics represented by mobile genetic elements, such as bacteriophage and Staphylococcal pathogenicity islands, and non-synonymous mutations affecting antibiotic resistance and virulence. Hospitals in large cities were identified as hubs for the transmission of MRSA to regional health care centres. In addition, comparison of whole genome sequences revealed that at least 3

independent acquisitions of TSST-1 have occurred in CC30, but a single distinct clade of diverse community-associated CC30 strains was responsible for the TSS epidemic of the late 1970s, and for subsequent cases of TSS in the UK and USA.

Finally, whole genome sequencing was used as a tool for investigating a recent outbreak of legionellosis in Edinburgh. An unexpectedly high level of genomic diversity was identified among the outbreak strains, with respect to core genome polymorphisms, and accessory genome content. The data indicate that affected individuals may be infected with heterogeneous strains. The findings highlight the complexities in identifying environmental sources and suggest possible differences in pathogenic potential among isolates from a single outbreak.

Taken together, the findings demonstrate applications of bacterial genome sequencing leading to enhanced understanding of bacterial pathogen evolution, emergence, and transmission, which may ultimately inform appropriate infection control measures.

1. Introduction

1.1. Bacterial diseases of humans

Bacteria have been the causative agents of some of the most widespread and fatal epidemics throughout human history. At the beginning of the 20th century, pneumonia, tuberculosis, and enteritis were the 3 leading causes of death in the United States, and along with diphtheria accounted for one third of all deaths (CDC, 1999). A century later, the mortality rate attributed to bacterial pathogens has significantly declined, with the theory of epidemiologic transition attributing the decline to a combination of socioeconomic factors (increased standards of living, hygiene practices, nutrition), and medical and public health interventions (antibiotic therapies, vaccination strategies, water sanitation) (Omran, 1971).

Despite the decline of the bacterial disease burden in developed nation populations, there have been numerous examples in the past 4 decades of the emergence of novel bacterial pathogens, and re-emergences of known bacteria with an increased capacity to cause disease (Table 1.1). Faced with the growing problem of antibiotic resistant bacteria (Davies et al., 2010; Kumarasamy et al., 2010), and the lack of development of new antibiotic therapies, it seems unlikely that the factors contributing to the decline in bacterial disease in the 20th century will have the same impact again. Consequently, an increased understanding of the processes that lead to the emergence of virulence, antibiotic resistance and the drivers of pathogen spread may help inform control programs through targeted interventions.

This thesis will explore the evolution and epidemiology of bacterial pathogens through the application of whole genome sequencing (WGS) of populations of *Staphylococcus aureus* and *Legionella pneumophila*.

Table 1.1 Selected emerging or re-emerging bacterial pathogens of humans.

Bacterium	Disease	Year first described	Reference
Bartonella henselae	Bartonellosis	1992	(Regnery et al., 1992)
Borrelia burgdorferi	Lyme disease	1982	(Burgdorfer et al., 1982)
Clostridium difficile	Pseudomembranous colitis	1978	(Larson et al., 1978)
Escherichia coli O157:H7	Haemorrhagic colitis	1982	(Riley et al., 1983)
Helicobacter pylori	Gastric Ulcer	1982	(Marshall et al., 1984)
Legionella pneumophila	Legionellosis	1977	(McDade et al., 1977)
Listeria monocytogenes	Listeriosis	1981	(Schlech et al., 1983)
Staphylococcus aureus	Toxic shock syndrome	1982	(Altemeier et al., 1982)
Streptococcus pyogenes	Streptococcal toxic shock syndrome	1989	(Stevens et al., 1989)

1.2. *Staphylococcus aureus*

The Gram-positive bacterium *S. aureus* is an important human pathogen responsible for an array of diseases (Lowy, 1998). The microorganism is commonly isolated as a commensal from the anterior nares and skin, with nasal carriage being a major risk factor for subsequent invasive disease (Davis et al., 2004; Wertheim et al., 2004). A review of cross-sectional studies showed the mean prevalence of *S. aureus* nasal carriage to be 37.2 % in the general population (Kluytmans et al., 1997). Longitudinal studies have demonstrated that approximately 20 % to 30 % of individuals are persistent nasal carriers of *S. aureus*, with 30 % carrying *S. aureus* intermittently, and the remainder of the population being non-carriers (Frank et al., 2010). However, an important recent study found that after experimental nasal colonisation with *S. aureus*, the elimination kinetics and antibody responses of individuals identified as either intermittent- or non-carriers were highly similar (Van Belkum et al., 2009). Swabs from individuals classed as persistent carriers yielded significantly higher CFU per sample, and were culture positive for longer than swabs from intermittent- or non-carriers. Taken together, these data suggest that human *S. aureus* nasal carriers fall into one of two classes, those who are persistently colonised, and others who may carry *S. aureus* given particular selective pressures (Van Belkum et al., 2009).

Invasive disease occurs when the protective barriers of the skin or mucosal membranes are compromised, typically at surgical or vascular access sites in the hospital setting. Infection is often localised to the lesion site, but can develop into systemic disease, particularly if the host is immunosuppressed (Lindsay, 2008;

Mourvillier et al., 2004). Clinical presentation ranges from relatively minor skin and soft tissue infections, to potentially fatal bacteraemia, infective endocarditis, toxic shock syndrome and necrotising pneumonia (Altemeier et al., 1982; Gillet et al., 2008; Mourvillier et al., 2004; Stryjewski et al., 2008; Wertheim et al., 2004).

1.2.1. Organisation and evolution of the *S. aureus* genome

At the time of writing there are 33 finished genome sequences of *S. aureus* available from the Genomes Online Database (GOLD) (Pagani et al., 2012), with over 8000 additional sequences identified in the NCBI sequence read archive (SRA) (*SRA Handbook*, 2010).

The *S. aureus* genome consists of a single chromosome with a GC % of approximately 33 %, ranging in size from 2.7 Mb encoding 2,545 CDSs (Sass et al., 2012) to 3.1 Mb encoding 2,977 CDSs (Holden et al., 2010). A comparative genomic analysis of 17 strains identified 2,266 conserved CDSs (approximately 75 % of CDSs) that make up the core genome of *S. aureus*, and 899 CDSs that were variably present and represent the accessory genome (Boissy et al., 2011). There is a high level of synteny maintained across the core genome of the *S. aureus* species (Lindsay et al., 2006) and, with the exception of strains from clonal complex 75, orthologous genes in the core genome typically display less than 2 % divergence in their nucleotide sequence (Holt et al., 2011).

Unlike bacterial genera such as the Streptococci and the Neisseriaceae, *S. aureus* is a clonal organism and genetic variation evolves primarily by point mutation, rather

than recombination (Feil et al., 2003). However, recombination is known to play an important role in the long term evolution of *S. aureus* (Robinson et al., 2004). The MRSA ST239 clone is prevalent in hospitals globally (Aires de Sousa et al., 1998, 2001; Alp et al., 2009; Feil et al., 2008) and has arisen as a result of a large scale chromosomal recombination event between genomes from the unrelated clonal complexes (CC) CC30 and CC8 (Robinson et al., 2004). In addition, recombination is known to impact on the diversification of a number of virulence determinants, including *agr* (Robinson et al., 2005b), staphylocoagulase (Watanabe et al., 2009), and staphylococcal enterotoxin-like toxin X (Wilson et al., 2011).

The accessory genome consists of a diverse collection of non-essential genes that can confer selective advantages in specific niches and can comprise approximately 20 % of a *S. aureus* genome (Fitzgerald et al., 2001). Within lineages, frequent horizontal transfer of mobile genetic elements (MGE) has been reported (Lindsay et al., 2006; McCarthy et al., 2010). Strikingly, the majority of staphylococcal toxins causing toxinoses are carried on MGEs. MGEs integrated into the *S. aureus* chromosome include bacteriophages, staphylococcal pathogenicity islands (SaPI), plasmids, transposons, and staphylococcal cassette chromosomes (SCC) (Novick, 2003).

1.2.1.1. Staphylococcal pathogenicity islands

SaPIs are a family of MGEs with 6 defined integration sites in the *S. aureus* genome, and are typically between 14 and 17 kb in size (Novick et al., 2007). SaPIs do not encode the capacity for horizontal transfer within their genome, and are closely associated with specific phage that mediate their excision from the *S. aureus* genome and replication (Tormo-Más et al., 2010). SaPIs maintain a core genome for the control of their replication cycle, including an integrase that determines the chromosomal attachment site, and a number of replication regulators (Novick et al., 2010). Importantly, many SaPIs encode superantigens, with toxic shock syndrome toxin (TSST-1) and enterotoxins B and C, superantigens associated with toxic shock syndrome (TSS) and food poisoning, being found exclusively on pathogenicity islands (Novick, 2003).

1.2.1.2. Bacteriophages

Bacterial prophages play an important role in the evolution of the *S. aureus* genome, and have the ability to disrupt host genes, while mediating the horizontal transfer of genes encoding virulence factors by transduction. Phage genomes are typically highly mosaic, with recombination between phage a common occurrence (Kwan et al., 2005). Phage can carry factors beneficial for survival in a novel host (Lowder et al., 2009), virulence factors for establishing and maintaining infection (Coleman et al., 1989; Diep et al., 2010), and immune evasion determinants (Van Wamel et al., 2006).

1.2.1.3. Staphylococcal cassette chromosomes

Staphylococcal cassette chromosomes that carry the *mecA* or *mecC* gene (SCC*mec*) are large mobile elements that encode PBP2a, a penicillin binding protein with low affinity for β-lactam antibiotics (Ubukata et al., 1989). SCC*mec* elements integrate into the *S. aureus* chromosome downstream of the *orfX* locus, and can be subtyped based on the nucleotide sequence of the *mec* gene complex, and of the *ccr* genes (Hanssen et al., 2006). The *ccr* loci encode site-specific chromosomal recombinases that mediate transfer of SCC*mec* by controlling the excision and insertion of the SCC*mec* element (Hanssen et al., 2006). To date, 11 SCC*mec* types conferring resistance to different classes of antibiotics have been described (García-Álvarez et al., 2011). SCC*mec* types are defined by the combination of the *mec* gene complex and the *ccr* gene complex carried by the element, and can be further subtyped based on the flanking nucleotide sequence of the *mec* and *ccr* complexes (Zhang et al., 2005). SCC*mec* I, II and III range in size between 34 – 67 kb and are predominantly associated with the nosocomial environment, in part because they encode resistance to multiple classes of antibiotics (Chambers et al., 2009; Collins et al., 2010). SCC*mec* types IV and V are smaller, approximately 21 – 28 kb, and are not restricted to the hospital environment, and are carried by *S. aureus* strains that are capable of causing disease in the community (Holden et al., 2013; Ma et al., 2002). It has been demonstrated that strains carrying the type II SCC*mec* have reduced levels of virulence factor expression and fitness relative to strains harbouring type IV SCC*mec* (Collins et al., 2010; Rudkin et al., 2012).

Understanding the composition of the core and accessory genome is essential in order to identify the evolutionary processes that may lead to the emergence of *S. aureus* with phenotypic traits that may cause an increased burden on public health, such as antibiotic resistance and increased virulence.

1.2.2. Antibiotic resistant *S. aureus*

The growing prevalence of resistance to all classes of antimicrobials is a major public health concern, complicating efforts to prevent and treat infections (Aires de Sousa et al., 2003; Begier et al., 2004; DeLeo et al., 2011; Harris et al., 2010; Holden et al., 2013). *S. aureus* readily acquires antimicrobial resistance via gene mutation or by the horizontal acquisition of resistance determinants from other bacteria (Holden et al., 2004). For example, single nucleotide polymorphisms in *fusA*, *rpoB*, *dfrB* and *gyrA* mediate resistance to fusidic acid, rifampin, trimethoprim and fluoroquinolones, respectively (Besier et al., 2003; Tanaka et al., 2000; Vickers et al., 2009; Wichelhaus et al., 2002).

The first methicillin resistant *S. aureus* (MRSA) strains were described in 1961, 6 months after the introduction of methicillin as an anti-staphylococcal agent (Jevons, 1961), but the incidence was low and these strains did not become endemic in Europe or the United States (Bran et al. 1972, Rosdahl et al., 1991; Rosendal et al., 1977). A second wave of MRSA clones emerged globally throughout the 1980s and 1990s and spread rapidly between hospitals, with reports of multidrug resistant strains in the UK, Ireland, Australia and the United States (Ayliffe, 1997; Cox et al., 1995; Marples et al., 1992). Currently, MRSA is disseminated on a global scale in

both nosocomial and community settings, with methicillin resistance being recorded in 40 % to 60 % of hospital-acquired *S. aureus* infections from Europe, Japan, and the United States (Grundmann et al., 2006; Tiemersma et al., 2004). Resistance is mediated by the expression of a penicillin-binding protein with low affinity for β-lactam antibiotics (PBP2a) (Fischetti et al., 2006), encoded by the *mecA* gene carried on the staphylococcal cassette chromosome *mec* (SCC*mec*), a large (21-67 kb) mobile genetic element (Hanssen et al., 2006).

Multidrug resistant strains can be treated using vancomycin as a last resort, although vancomycin resistant strains have been sporadically reported since the late 1990s (Hiramatsu et al., 1997). High-level resistance to vancomycin is mediated through expression of the VanA operon, carried on the *Tn1546* transposon, which was acquired through horizontal transfer from enterococci (Kos et al., 2012; Sievert et al., 2008). Intermediate resistance to vancomycin has been linked to variants of numerous loci, including *vraSR*, a 2-component system that controls the expression of a number of genes involved in peptidoglycan cross-linking, and thickening of the bacterial cell wall (Gardete et al., 2006; Mwangi et al., 2007).

S. aureus is capable of causing disease in both hospital and community niches, and emerging strains are increasingly capable of causing infections in both environments (Espadinha et al., 2013). An understanding of the factors driving the emergence and maintenance of clones is essential in order to design and implement effective control measures.

1.2.3. Hospital-associated MRSA

Hospital-associated MRSA (HA-MRSA) is a leading cause of nosocomial infection, and significantly increases morbidity and mortality rates of affected inpatients. Since the 1990s, multidrug-resistant MRSA has been endemic in UK hospitals, resulting in government interventions to increase hygiene standards and bring the prevalence of MRSA infections down. Throughout the late 1990s and early 2000s *S. aureus* was the most common cause of nosocomial-acquired infection, with 2 % of all new hospital admissions in the United States resulted in the patient suffering an MRSA infection (Jones, 2003), while this figure was 1.3 % in the United Kingdom (Smyth et al., 2008). Although HA-MRSA rates have fallen, the cost of MRSA to the European economy is estimated to be €380,000,000 per year (Köck et al., 2010). Estimates from the United States attribute 477,927 hospitalisations to *S. aureus*, resulting in death in over 11,000 cases in the years 1999-2005 (Klein et al., 2007). More recent figures from the United Kingdom for 2010/11 report the rate of MRSA bacteraemia as 3.2/100,000 bed days, while the rate of methicillin sensitive *S. aureus* (MSSA) bacteraemia is 24.6/100,000 bed days (Health Protection Agency, 2012a). *S. aureus* is the second most common cause of post-operative infection in the United Kingdom, accounting for 27 % of surgical site infections in 2010/11 (Health Protection Agency, 2012b). Globally, prevalence of HA-MRSA has been reported at over 70 % in regions of Asia, over 50 % in the Americas and regions of Europe, while rates in Northern Europe can be as low as 5 % (Stefani et al., 2012).

Among HA-MRSA, 6 clonal complexes (CC5, 8, 22, 30, 45 and 239) dominate the population, with differences in prevalence associated with geographic location

(Chambers et al., 2009; Lindsay, 2013). CC5 with SCC*mec* types II and IV, and SCC*mec* type IV ST8 strains predominate in the USA, in southeast Asia and South America, the dominant clone is ST239. Within Europe, there are a number of HA-MRSA lineages, with CC5 predominating in central regions (Schulte et al., 2012). In the UK, the epidemic clones EMRSA-15 and EMRSA-16 predominate among the circulating strains and have been responsible for over 95 % of cases of HA-MRSA since the mid-1990s, although the prevalence of EMRSA-16 has been rapidly declining (Ellington et al., 2010; Johnson et al., 2005; Murchan et al., 2004; Wyllie et al., 2011). The decline in rates of EMRSA-16 have been hypothesised to be due to changes in hygiene management, antibiotic prescription patterns, and phage epidemics (Ellington et al., 2010; Knight et al., 2012; Wyllie et al., 2011)

1.2.4. Community-associated MRSA

In contrast to HA-MRSA, which primarily affects individuals predisposed to infection, community-acquired infections (CA-MRSA) commonly occur in otherwise healthy individuals (Herold et al., 1998; Söderquist et al., 2006). Outbreaks of the community-associated USA300 clone have been reported in athletics teams, inmates at correctional facilities, military recruits, and men who have sex with men (Diep et al., 2008; Maree et al., 2007). CA-MRSA lineages typically demonstrate increased virulence in comparison to HA-MRSA, and although commonly associated with severe skin and soft tissue infections, can occasionally cause life-threatening necrotising pneumonia, and other severe infections (Boyle-Vavra et al., 2007; Collins et al., 2010). Colonisation sites include the throat, axilla, groin, perirectal

area and non-intact skin sites, but nasal colonisation does not appear to be a risk factor for CA-MRSA infection (DeLeo et al., 2010).

Epidemics of multidrug resistant strains, and the emergence of new pathogenic clones in the community, highlight the importance of understanding the molecular basis for resistance phenotypes and the evolutionary events that lead to their emergence. Increasingly, the boundaries between the hospital and the community are becoming blurred, with virulent CA-MRSA lineages causing severe infections within a healthcare setting, and HA-MRSA circulating in community settings (Espadinha et al., 2013; Sherwood et al., 2013; Torres-Sangiao et al., 2011). By understanding the dynamics of transmission between hospital and community settings, risk factors may be identified leading to the rational design of infection control strategies.

1.2.5. *S. aureus* CC30 is a clone of global health importance

Of the major globally disseminated lineages, *S. aureus* CC30 has had a particularly profound impact on human health over a period of 6 decades, giving rise to 3 pandemic lineages and the strains responsible for the TSS epidemic (Alesana-Slater et al., 2011; Altemeier et al., 1982; Blair et al., 1960; Collignon et al., 1998; DeLeo et al., 2011; Donahue et al., 1966; Fitzgerald et al., 2001; Nahmias et al., 1961; Robinson et al., 2005a).

In the 1950s a clone of MSSA spread rapidly from hospitals, causing significant skin and soft tissue infections in the community. Designated phage type 80/81, the clone

was characterised by resistance to penicillin, and production of the Panton-Valentine leukocidin (PVL) toxin (Donahue et al., 1966; Nahmias et al., 1961). The prevalence of phage type 80/81 declined throughout the 1960s and 1970s with the introduction of new antibiotics (Blair et al., 1960; Goldie et al., 1971).

Emerging in the 1990s in remote Pacific Island populations, the Southwest Pacific clone (SWP) is a contemporary PVL positive MRSA clone. Like the phage type 80/81 clone, the SWP clone is community-associated and spread globally causing skin and soft tissue infections of otherwise healthy individuals (Alesana-Slater et al., 2011; Collignon et al., 1998).

In contrast to phage type 80/ 81 and SWP, the EMRSA-16 (ST36) clone is restricted to the hospital setting and has reduced virulence due in part to low levels of expression of cytolytic toxins (DeLeo et al., 2011). The clone has been endemic in UK hospitals for over 20 years, although the prevalence has declined rapidly in recent years, partly as a result of changes in infection control and antibiotic prescribing regimens (Ellington et al., 2010; Knight et al., 2012; Wyllie et al., 2011). EMRSA-16 has been reported less commonly throughout Europe, Southeast Asia, South Africa, Australia, and North America (Carleton et al., 2004; Ho et al., 2009; Jansen van Rensburg et al., 2011; Pérez-Roth et al., 2004; Robinson et al., 2005a; Scicluna et al., 2010).

Previously, the SWP clone was proposed to be a re-emergence of the phage type 80/81 clone after acquisition of SCC*mec*-IV, and this hypothesis was supported by a

combination of MLST and *spa* typing, along with comparison of the surface associated and virulence proteins carried by each clone (Diep et al., 2006; Robinson et al., 2005a). The same model of CC30 evolution proposed that the EMRSA-16 clone evolved, through the acquisition of SCC*mec*-II, in parallel with the community-associated lineages from a PVL negative ancestor, leading to the emergence of EMRSA-16. However, a recent study based on phylogenetic reconstruction and comparative genome sequence analysis of 9 CC30 isolates demonstrated that the SWP clone is not a direct descendent of the phage 80/81 clone, but that each clone has evolved from a single ancestral strain that gave rise to all three CC30 pandemics, and the strains responsible for the TSS epidemic (DeLeo et al., 2011). While DeLeo and colleagues demonstrated that the EMRSA-16 lineage have mutations in *agr* and *hla* genes that lead to reduced virulence in mouse models of infection, these mutations are also present in isolates from TSS infections in the community, indicating that they are not sufficient for restriction to the hospital environment (DeLeo et al., 2011).

An understanding of the basis for the niche adaptation of each CC30 lineage is currently lacking, and comparative genomic analysis of CC30 strains could yield important insights into the molecular processes underlying the emergence of pandemic lineages.

The repeated global emergence of *S. aureus* clones that have a major public health and economic costs highlights the need to understand the molecular correlates driving emergence and expansion.

1.3. *Legionella pneumophila*

L. pneumophila is a Gram-negative, aerobic, non-capsulated bacillus (Heuner et al., 2008). Over 50 *Legionella* species have been identified, over half of which have been associated with human disease (Fields et al., 2002). *L. pneumophila* is a ubiquitous intracellular pathogen of environmental protozoa that inhabits soil and freshwater reservoirs (Rowbotham, 1980).

L. pneumophila was first isolated in 1976 following an outbreak of pneumonia among attendees of a convention of the American Legion at a hotel in Philadelphia, in which 182 people were infected and 29 died (Fraser et al., 1977; McDade et al., 1977). *L. pneumophila* can cause a spectrum of diseases, termed legionellosis, although some individuals can seroconvert without displaying any symptoms (Rudbeck et al., 2009). Pathogenesis is mediated through the intracellular replication of *L. pneumophila* in human phagocytes and the Icm/Dot type IVB secretion system delivering effector proteins to the host cell (Luo et al., 2004; Nagai et al., 2002; Zusman et al., 2003). Fusion of the phagosome and lysosome is prevented, inhibiting vacuole acidification and allowing proliferation of the bacteria leading to cell death by apoptosis, or necrosis mediated by pore-forming toxins (Gao et al., 1999; Sadosky et al., 1993; Tateda et al., 2003).

Legionellosis only occurs when an individual inhales an aerosol containing *L. pneumophila* and there have been no reported cases of onward transmission from infected individuals, suggesting that the environment is exclusively responsible for the infection source (Katz et al., 1982). The risk of aerosolisation increases with

high levels of *L. pneumophila* in water reservoirs, which can occur if water reservoirs are not maintained to a high standard (Rogers et al., 1994). *L. pneumophila* can persist in biofilms, and is more readily cultured from biofilm samples than from flowing water, suggesting that biofilm removal is an important measure for control of *L. pneumophila* levels in water reservoirs (Rogers et al., 1994). Several man-made water systems have been found to contain *L. pneumophila*, including cooling towers, hot tubs, and nebulizers (Mastro et al., 1991; Nguyen et al., 2006; Orrison et al., 1981; Thomas et al., 1993).

Increased temperature and increased vapour pressure are significantly associated with aerosolisation (Conza et al., 2013), while incidence of legionellosis is greater during the summer season (Fisman et al., 2005; Li et al., 2002) and during periods of increased humidity (Fisman et al., 2005). Distance from the environmental source (Nguyen et al., 2006), wind direction (White et al., 2013), and the *L. pneumophila* dose that an individual is exposed to (Brown et al., 1999) are important risk factors for infection.

The non-fatal Pontiac fever form of legionellosis presents with symptoms similar to acute influenza (Fraser et al., 1979), while the more severe Legionnaires' disease (LD) presents as a pneumonia and has a case fatality rate of up to 34 % (Benin et al., 2002). Clinical manifestations are variable between patients, and are of limited use for diagnosing Legionnaires' disease. Therefore, a diagnosis of legionellosis can only be made based on the results of appropriate laboratory tests, such as the urinary antigen test, in patients presenting with pneumonia.

Human susceptibility to *L. pneumophila* infection is heterogeneous within the population, with smoking, older age, diabetes, malignancy, AIDS, and end stage renal disease being significantly associated with developing LD (Marston et al., 1994). Mortality as a result of LD has a significant association with older age, male gender, immunosuppression, end stage renal disease, and nosocomial acquisition of disease (Marston et al., 1994).

The incidence rate of legionellosis is variable between countries. In Scotland the annual incidence rate of legionellosis was 4.8/1,000,000 in 2009 and 3.4/1,000,000 in 2010. By comparison, the incidence rate across Europe is considerably higher, averaging 11.8/1,000,000 during 2007-08 and 11.2/1,000,000 in 2009 (Potts et al., 2011).

1.3.1. Genome organisation of *L. pneumophila*

The genome of *L. pneumophila* is approximately 3.4 Mb in length, representing approximately 3,000 CDSs (Cazalet et al., 2004; Chien et al., 2004a). Analysis of the available *L. pneumophila* whole genome sequences indicates that there has been substantial horizontal gene transfer among *L. pneumophila* strains (D'Auria et al., 2010; Glöckner et al., 2008), and moreover several homologs of eukaryotic genes have also been identified in their genome (Cazalet et al., 2004; Chien et al., 2004b). A previous analysis of 5 whole genome sequences of *L. pneumophila* demonstrated that 67 % of the total identified CDSs are conserved in all strains, while 33 % are

part of a dispensable accessory genome (D'Auria et al., 2010), indicating a high level of genomic plasticity within the species.

Type IV secretion systems (T4SS) are complexes of proteins that are required for conjugation-mediated genetic exchange, and can deliver effector proteins into host cells during infection (Fronzes et al., 2009); they are variably present among the sequenced *L. pneumophila* strains (D'Auria et al., 2010; Glöckner et al., 2008). Three families of T4SS have been identified, including T4ASS, T4BSS and the genomic island-associated T4SS (GI-T4SS) (Juhas et al., 2007). Among the sequenced *L. pnueumophila* strains, 3 T4ASSs have been described; Lvh, Trb-1 and Trb-2 (Glöckner et al., 2008). The Lvh T4ASS has a role in host-cell entry and intracellular replication (Bandyopadhyay et al., 2007), and contributes to establishment of infection at lower temperatures (Ridenour et al., 2003). The roles of the Trb-1 and Trb-2 T4ASS remain to be determined, but a recent study showed that Trb-1 is not required for intracellular replication of *L. pneumophila* (Lautner et al., 2013).

The Dot/Icm system is encoded by a T4BSS and is essential for intracellular replication by translocating effector proteins into the host cell cytoplasm (Luo et al., 2004; Nagai et al., 2002; Zusman et al., 2003). The functions of many of the effector proteins are unclear, but a large number show significant homology to eukaryotic proteins (Brüggemann et al., 2006; Cazalet et al., 2004; Gomez-Valero et al., 2011), which may allow *L. pneumophila* to modulate the host response through structural mimicry of host components (Lomma et al., 2009).

1.3.2. Investigation of outbreaks of legionellosis

Due to the threat posed to public health by *L. pneumophila* infection, determining the source of a legionellosis outbreak is critical in order for appropriate control measures to be implemented. Sources are identified through a combination of typing and epidemiological methods. There are 16 identified serogroups (Sg) of *L. pneumophila*, Sg 1, 3 and 6 being most associated with human disease, with Sg 1 responsible for approximately 80 % of reported legionellosis in humans (Fields et al., 2002). *L. pneumophila* isolates of Sg 1, 4 and 5 can be assigned a monoclonal antibody (mAb) subgroup by indirect immunofluorescence or ELISA methods (Lück et al., 2013). Further, a DNA sequence based typing (SBT) scheme using the nucleotide sequence of 7 loci under diversifying selection can be used to allocate an unambiguous sequence type (ST) to *L. pneumophila* isolates (Gaia et al., 2005; Ratzow et al., 2007).

Spatial analysis of cases, combined with knowledge of meteorological conditions can suggest a likely water reservoir as a source (McCormick et al., 2012; White et al., 2013), which can then be confirmed if the Sg, mAb type, and SBT of *L. pneumophila* from that source are concordant with the isolates from clinical cases. The application of whole genome analysis to *L. pneumophila* outbreaks has the potential to greatly improve the understanding of both the epidemiology and the biology of a pathogen with an important impact on public health.

Recently, whole genome sequencing has been used in an attempt to determine the source of an outbreak of *L. pneumophila* in Hampshire, UK, which resulted in a total

of 26 cases of LD over a 3 week period in 2003. Traditional typing methods implicated 2 cooling towers on a single site as the likely source of the outbreak (Reuter et al., 2013). The retrospective application of whole genome sequencing of 2 outbreak strains and 4 environmental strains isolated at the time of the outbreak facilitated the identification of a close phylogenetic relationship between isolates from the cooling towers and isolates from patients. However, an isolate obtained from a domestic spa pool also clustered with the outbreak strains, meaning this site cannot be ruled out as a source of the outbreak. An isolate from a cooling tower on a separate site was phylogenetically distant to the outbreak strains and had previously been eliminated as being the likely source due to its distinct monoclonal antibody subgroup type (Reuter et al., 2013).

The high case fatality rate associated with LD means that outbreaks must be rapidly contained through the identification and treatment of environmental sources. The impact of whole genome sequencing on the understanding of *L. pneumophila* outbreaks remains to be thoroughly assessed, but it may be able to provide useful information for the design of public health control programs.

1.4. Bacterial population genetic methods

Population genetics is the study of allele frequency within a population, and of the change in genetic composition within a population through processes including genetic drift, recombination and natural selection. Therefore, population genetic analysis can give an insight into the evolutionary history of an organism and inform on correlates of niche adaptation and population divergence (Okasha, 2012).

Traditional typing techniques employed by epidemiologists studying bacterial pathogen populations suffer from a number of limitations. Genetic fingerprinting methods such as pulsed field gel electrophoresis (PFGE) and enterobacterial repetitive intergenic consensus polymerase chain reaction (ERIC-PCR) are difficult to standardise between laboratories, are prone to subjective interpretation, and may not be applicable to some bacterial species due to degradation of genomic DNA by endonucleases during electrophoresis (Sampaio et al., 2006; van Belkum et al., 1998).

Typing methods based on DNA sequence offer an unambiguous means of typing bacterial strains, but are also not without limitations. For example, staphylococcal protein A (*spa*) typing method is based on differences in tandem repeats in the major *Staphylococcus aureus* surface protein A, SpA, and is inexpensive and reproducible between laboratories (Koreen et al., 2004). However, *spa* typing suffers from a lack of discriminatory power (Cookson et al., 2007) and incorrect clonal complex assignments have been reported due to homoplasies (Schouls et al., 2009).

Multilocus sequence typing (MLST) assigns an isolate to a clonal complex based on the nucleotide sequence of internal fragments of 7 housekeeping genes found in the core genome. Each sequenced fragment is assigned an allele number, and the combination of alleles at the seven housekeeping loci defines the ST of the isolate (Maiden et al., 1998). MLST has advantages over other methods, as the level of sequence diversity at the housekeeping loci within a bacterial species is sufficient for billions of potential sequence types to be assigned, and at the time of writing there are over 2,700 assigned STs of *S. aureus*. MLST has been widely adopted and applied to the study of the evolution and epidemiology of major human pathogens including *S. aureus* (Enright et al., 2000), *Helicobacter pylori* (Wirth et al., 2004), and *Streptococcus pneumoniae* (Enright et al., 1998). It is particularly useful for examining species-wide population structure, but is limited when examining diversification over short timescales, such as in outbreak investigations, as there is unlikely to be sufficient accrued sequence variation among the MLST loci to determine likely sources or transmission pathways.

Furthermore, strains of the same ST often have remarkably different phenotypes related to virulence, resistance, and tissue/host tropism. Specialist phenotypes often arise through expression of genes encoded by accessory elements, which can be rapidly acquired by or lost from a population. Therefore 2 closely related isolates, with conserved core genomes, as assessed through traditional typing methods, might be phenotypically very divergent. WGS offers a high-resolution view of the whole genome of an isolate, and important differences in accessory genome content can be identified. As an example, the isolates MW2 and MSSA476 are both typed as *S.*

aureus ST1 by MLST. Both isolates belong to a highly virulent clone of *S. aureus* and infected immunocompenent individuals in the community, but were associated with very different pathologies (Baba et al., 2002; Holden et al., 2004). MW2 caused a fatal bacteraemia, whereas MSSA476 was responsible for an osteomyelitis from which the patient made a full recovery. These differences in disease outcome could be as a result of differing host factors, or may be attributable to five large scale genome changes leading to increased virulence of MW2. Among these large scale changes are the acquisition by MW2 of SCC*mec*, leading to multidrug resistance, and a phage encoding the PVL loci, which is a marker for community-associated clones of *S. aureus* with elevated levels of virulence (Diep et al., 2010).

1.5. Genomic epidemiology

The ability to rapidly sequence large numbers of bacterial genomes is revolutionising the study of bacterial population genomics, evolution, and epidemiology. While previously WGS of bacterial isolates represented a picture of a single organism at a given point in time, it is now possible to study the extent of variation within a pathogen population through space and time.

These advancements have been made possible through the rapid pace of development of sequencing technologies that have seen a shift away from using conventional Sanger sequencing for WGS of bacterial pathogens (Metzker, 2010). High-throughput technologies such as Illumina's HiSeq, Roche's 454 GS-FLX+, and Pacific Biosystems RS platforms have the capacity to generate enormous volumes of sequence data rapidly and economically (Loman et al., 2012a). Additionally, the

availability of high-throughput benchtop sequencers such as the Illumina MiSeq, Roche 454 Junior, and Life Biosciences Ion PGM are offering the ability to rapidly sequence pathogens to clinical and basic science research labs outside of specialist sequencing centres (Eyre et al., 2012; Loman et al., 2012b). Increasingly, modern infectious disease epidemiology is making use of the higher level of resolution afforded by WGS to enhance the understanding of epidemiology over varying scales. It is now possible to study the evolution of a pathogen during infection of a single patient, reconstruct transmission pathways during outbreaks, and to elucidate the molecular mechanisms underpinning the emergence of successful epidemic or pandemic clones.

1.5.1. Within-host evolution

The dynamics between pathogen and host during infection determine disease progression. Our understanding of bacterial evolution during infection has lagged behind that of infecting virus populations due to the lower mutation rate and the prohibitive costs associated with WGS of bacteria.

A number of studies of the genome sequences of HIV isolated from a single patient have shown that it is possible to draw clinically relevant conclusions about the rate of disease progression from the mutation rate of the virus (Lemey et al., 2007; Mikhail et al., 2005). Specifically, Lemey and colleagues identified that increased synonymous substitution rates were associated with shorter virus generation times and more rapid disease progression (Lemey et al., 2007). This finding can directly

impact on clinical practice, by identifying patients at high risk of progressing from HIV to AIDS, allowing treatment to be tailored appropriately.

Recently, WGS studies of bacterial populations within a single host have offered valuable insights into the evolutionary processes during the course of colonisation and infection (Bryant et al., 2013; Golubchik et al., 2013; Lieberman et al., 2011; Mwangi et al., 2007; Smith et al., 2006; Young et al., 2012). Within a single host, 'clouds' of genetic heterogeneity have been described for pathogens including *S. aureus, Mycobacterium tuberculosis, Mycobacterium abscessus,* and *Burkholderia dolosa,* due to diversification of the bacterial population during carriage or infection (Bryant et al., 2013; Ford et al., 2011; Harris et al., 2013; Lieberman et al., 2011; Young et al., 2012). Over time, as colonisation or infection progresses, random mutations become fixed due to host-specific selective pressures including coinfection with other microorganisms, the host immune system, and antimicrobial therapies (Lieberman et al., 2011; Mwangi et al., 2007; Smith et al., 2006; Young et al., 2012; Zdziarski et al., 2010). In particular, a common phenomenon of persistent bacterial infection is the emergence of reduced virulence, antibiotic resistant bacteria associated with chronic infection and low mortality (Gao et al., 2010; Smith et al., 2006). For example, a comparative genomic study of *Pseudomonas aeruginosa* during chronic infection of a cystic fibrosis (CF) patient revealed a strong signal of genome-wide positive selection indicated by a dN/dS ratio of 1.4 (Smith et al., 2006). Of note, about 25 % of all mutations acquired over 96 months resulted in pseudogene formation through nonsense mutations leading to premature stop codons, frameshifts, transposon insertions or gene deletion events. Importantly, expression

of numerous virulence factors was selected against, reflecting a shift from acute infection towards an enhanced ability to cause chronic inflammation (Smith et al., 2006).

Similarly, loss of virulence factor production, associated with evolution towards commensalism has been described in multiple independent patients undergoing therapeutic bladder colonisation with the prototype asymptomatic bacteriuria strain *E. coli* 83972 for the treatment of recurrent urinary tract infections (Zdziarski et al., 2010). Sequencing of isolates recovered from the patients at sequential time points post-inoculation revealed numerous mutations relative to the progenitor strain, with an overall trend to reduced virulence and adaptation to oxidative stresses (Zdziarski et al., 2010). A number of mutations were unique to each host, and were recurrently observed in isolates obtained after repeat inoculation events in each patient, suggesting that the unique niche of each host strongly influenced the microevolution of the inoculated strain (Zdziarski et al., 2010). These data suggest that pathogens evolve differently in individuals, implying that personalised therapeutic regimes could be effective for the treatment of bacterial infections.

Conversely, convergent evolution in genes associated with antibiotic resistance and virulence has been observed during infection of multiple patients with *P. aeruginosa* and *B. dolosa* (Lieberman et al., 2011; Wong et al., 2011), and an understanding of the bacterial loci under selection in multiple patients can inform on treatment regimes and suggest targets for novel therapeutics.

Acquisition of antibiotic resistance during infection, and the underlying molecular mechanism, has been demonstrated in *S. aureus* through the use of WGS of sequential isolates obtained from a patient undergoing antibiotic therapy for infective endocarditis (Mwangi et al., 2007). Comparison of isolates obtained pre- and post-antimicrobial therapy with rifampin, imipenem, and vancomycin revealed a total of 35 point mutations in 31 loci, including 6 loci that had previously been implicated with increased resistance to a range of antibiotics (Mwangi et al., 2007). Importantly, the ability to track the *in vivo* evolution of sequential drug resistance has the potential to inform bespoke prescription regimens with the aim of minimising the emergence of antibiotic resistance traits.

As part of a long-term carriage study, Young et al. sequenced a total of 68 colonies of *S. aureus* isolated from 6 nasal swabs and a blood culture from a single patient with an ultimately fatal bacteraemia over a 13 month period (Young et al., 2012). 30 polymorphisms were identified over the course of the study, a sufficient level of diversity for 3 clusters to be identified in a maximum-likelihood phylogenetic reconstruction representing isolates from early nasal carriage, late nasal carriage, and invasive bloodstream infection, respectively. Strikingly, only 8 mutations separated late nasal carriage strains from strains isolated from the bloodstream, 50 % of which resulted in the premature truncation of protein coding genes. The authors speculate that 1 of these mutations, in an *AraC*-family transcriptional regulator may have resulted in an increased potential for pathogenicity, and played an important role in disease progression. The study clearly demonstrates the potential of whole genome sequencing for understanding the transition from commensal to pathogen, although

the feasibility of applying this methodology to larger cohorts remains to be determined.

It is important to understand the selective forces acting on a bacterial population within the host, and how they drive the evolutionary processes leading to adaptation to long-term colonisation and persistence, or from commensal to pathogen.

1.5.2. Using WGS to investigate outbreaks of bacterial infection

1.5.2.1. Identification of transmission events during localised outbreaks

Traditional typing methods for bacterial pathogens are often unable to discriminate between closely related strains, and are of limited use in tracing transmission routes during outbreaks. The high-resolution afforded by bacterial WGS means that the distribution of genetic polymorphisms within a population of microbes can be used to determine the spread of a pathogen in space and time (Eyre et al., 2012; Gardy et al., 2011; Köser et al., 2012; Lewis et al., 2010).

Analysis of the variation among bacterial genome sequences can identify potential transmission events between the hospital and community environments. The relatedness of *S. aureus* isolates obtained from 5 patients sharing a ward together for overlapping periods of 1-2 days, and one from a relative of one of the patients were assessed using *spa* and PFGE typing (Eyre et al., 2012). The *spa* and PFGE profiles were indistinguishable, and given the prolonged timescale of isolate collection these typing methods were insufficient to determine whether the outbreak isolates were derived from a point source, or circulating in the community. A single polymorphic

site was identified within the core genomes of the 6 isolates, suggesting a point source for the infection in the very recent past (Eyre et al., 2012).

Köser et al. investigated an outbreak of MRSA in a neonatal unit (Köser et al., 2012). MRSA isolates from 7 patients suspected of being part of the outbreak were sequenced, in addition to isolates from 7 patients that did not have an epidemiological link to the outbreak. Genomic analysis revealed all of the outbreak strains, as well as 3 non-outbreak strains to be ST22, and a total of 102 polymorphic sites separated the outbreak strains from non-outbreak strains. This high-resolution view allowed the authors to determine that the outbreak was self-contained, and that transmission of these strains was limited to the neonatal ward. In addition, a full profile of resistance and virulence determinants could be identified from the sequence data, further illustrating the utility of whole genome sequencing in a clinical setting (Köser et al., 2012).

Based on the results of multiple locus variable nucleotide tandem repeats analysis, a clonal population of *M. tuberculosis* was believed to be responsible for a sustained outbreak of tuberculosis in Canada (Gardy et al., 2011). WGS analysis revealed the existence of 2 extant circulating lineages, and integration of sequence data with social network information identified an at-risk population of crack cocaine users that maintained transmission of the outbreak. The combination of epidemiological and WGS data revealed the dynamics of the outbreak, information which could allow for targeted interventions to be designed, such as breaking transmission chains by treating highly connected individuals.

Often, the identification of outbreak isolates, and therefore the dynamics of the outbreak, can be obscured due to phenotypic differences between isolates that may have arisen due to horizontal gene transfer. Sequence analysis of the core genome offers an unambiguous method for determining the relatedness of isolates. For example, an outbreak of *S. aureus* on a single ward showed differences in the antibiotic sensitivity profiles between the isolates, despite identical *spa* and PFGE profiles, leading investigators to question whether multiple strains were circulating (Eyre et al., 2012). Application of WGS to 10 isolates from 8 affected individuals revealed a total of 4 variable sites in the core genomes of the isolates, providing strong evidence of an outbreak from a point source. Differences in resistance profiles in 2 isolates had initially caused confusion as to their origin, but this variation was found to be due to the carriage of resistance determinants on 2 plasmids (Eyre et al., 2012).

Through combining epidemiological data with analysis of bacterial population genome sequences, it is possible to infer the directionality of transmission in certain scenarios. Phylogenetic methods have long been applied to viral sequences, and have elucidated pathways of patient to patient transmission for HIV (Ou et al., 1992) and hepatitis C virus (Heinsen et al., 2000). One way to infer direction is based on patterns of genetic diversity within bacterial populations causing chronic infection. For example, analysis of *B. dolosa* genome sequences from 112 isolates from 14 CF patients over 16 years revealed that between patient transmission events could be inferred from the phylogenetic tree based on polymorphisms shared in isolates

obtained from different patients (Lieberman et al., 2011). A study of individuals with *M. abscessus* subspecies *massiliense* pulmonary infections identified greater diversity among isolates from a single patient, than among isolates from different patients (Bryant et al., 2013), strongly suggestive of between patient transmission of this pathogen. Additionally, isolates with high levels of macrolide resistance were obtained from 3 patients that had not previously received treatment with this class of antibiotic, providing further evidence for transmission between patients (Bryant et al., 2013).

Routine outbreak investigation can benefit greatly from the discriminatory power of WGS, and the integration of rapid sequencing of suspected outbreak strains can lead to changes in clinical practice and improved infection control. Harris et al. demonstrated that it is possible to infer transmission events within the hospital, and between hospital and community settings (Harris et al., 2013). A total of 34 polymorphisms between 24 isolates collected from a single neonatal ward confirmed that the isolates were part of a single prolonged outbreak. Interrogation of hospital records determined possible routes of transmission outside of the neonatal unit, with epidemiological links between parents of infants on the neonatal unit being established. However, the isolates had been sampled over a timeframe that included 3 periods in which the ward was determined to be free of MRSA. Persistence of the outbreak strain led the investigators to believe that an asymptomatically colonised healthcare worker was responsible for repeated transmission to patients. 20 isolates from a staff member positive for MRSA colonisation were sequenced and segregated into two groups. One group showed a close phylogenetic relationship to isolates

from before an MRSA-free period, and the other group most closely related to an isolate obtained after this time, strongly implicating the healthcare worker in transmission of MRSA over the course of the outbreak. Subsequent to decolonisation of the healthcare worker, no further cases of MRSA infection with the outbreak strain were identified, highlighting the role that whole genome sequencing can play in managing outbreaks. However, the possibility remains that the full level of diversity within each patient was not sampled, which could confound the findings of the study.

Taken together, these results suggest that WGS can aid the investigation of outbreaks at a local scale, and are able to identify transmission events within the hospital, in the community, and across the nosocomial/community barrier.

1.5.2.2. Identification of the emergence and spread of pandemic lineages

Bacterial populations accumulate polymorphisms over time as a result of genetic drift. Over appropriate timescales, the level of genetic heterogeneity is such that the population may be termed a measurably evolving population (MEP), allowing for the calibration of a molecular clock (Drummond et al., 2003; Gray et al., 2011; Holden et al., 2013). Utilising the accumulated levels of genetic variation within globally disseminated clonal lineages of bacteria, it is possible to reconstruct the demographic history of an organism, and potentially identify the geographic region and time period of clone emergence.

The *S. aureus* ST239 lineage is a globally disseminated multidrug resistant clone endemic in hospitals in Asia, and has been responsible for epidemics worldwide (Aires de Sousa et al., 2003; Alp et al., 2009; Smyth et al., 2010). Typing methods such as *spa* and PFGE typing provide a limited ability for discrimination of strains within a given location, as clonal expansion can lead to one strain dominating large geographic areas. Analysis of genome sequences from 63 isolates identified 4,310 polymorphisms within the core genome of these isolates. Subsequent phylogenetic reconstruction resulted in a robust tree that demonstrated geographic clustering of isolates within the phylogeny, facilitating identification of transmission pathways (Harris et al., 2010). Further analysis utilising a Bayesian framework estimated the median time of emergence for the clone as between 1957 and 1972, concordant with the introduction of methicillin as an anti-staphylococcal agent. Furthermore, incorporation of spatial data into the model shows frequent intercontinental transmission events throughout the 1990s (Gray et al., 2011).

Vibrio cholerae is the causative agent of cholera and has been responsible for multiple pandemics throughout recorded history. The current 7th pandemic, caused by the El Tor biotype, has given rise to 3 independent, but overlapping waves (Mutreja et al., 2011). WGS and phylogenetic reconstruction has identified several intercontinental transmission events, primarily emerging from a source population located in the Bay of Bengal (Mutreja et al., 2011). Comparative genomic analysis of *V. cholerae* isolated from patients during an epidemic following the 2010 Haitian earthquake showed these isolates are highly similar to strains circulating in Nepal,

implicating the movements of Nepali peacekeepers for transmission of cholera (Hendriksen et al., 2011).

The ability to identify global routes of bacterial pathogen transmission using WGS has the potential to inform on control measures to limit pandemic spread. For example, rapid identification of a source of a bacterial outbreak, such as the Haitian cholera epidemic, could help shape the public health response, and reduce the burden of disease.

1.5.3. Basis for the emergence of successful pathogenic lineages

Many bacteria capable of opportunistically causing a range of pathologies are ordinarily commensals and colonise their host without causing disease (Knapp et al., 1988; Tenaillon et al., 2010; Wertheim et al., 2005). Understanding the processes that lead to the selection for and emergence of successful pathogenic lineages that harbour virulence factors, immune evasion factors, and antibiotic resistance determinants could aid in developing treatment regimes that minimise the selective pressures for the emergence of virulent clones.

The 7th cholera pandemic has comprised 3 independent waves, with the emergence and spread of the latter 2 waves appearing to have been driven through the acquisition of the SXT antibiotic resistance element (Mutreja et al., 2011). The SXT element encodes resistance to several antibiotic classes and its initial acquisition appears to have occurred at the transition between the first and second pandemic waves. The success of the second, and subsequent, pandemics may in part be as a

result of the increased levels of resistance to the antibiotics commonly used to treat cholera during the first wave (Mutreja et al., 2011).

It has been shown that invasive *Salmonella* Typhimurium strains in sub-Saharan Africa belong to 2 highly conserved lineages within MLST-defined ST313. When compared in the context of a species-wide whole genome phylogeny to the highly diverse *Salmonella* Typhimurium strains circulating globally, it is clear that these 2 sub-Saharan African lineages represent distinct epidemics of invasive disease. Independent acquisition of a transposon encoding multidrug resistance genes has been postulated as a reason for the success of these 2 lineages. However, the rate of the molecular clock in these lineages reveals that the estimated time of clonal expansion of each lineage strongly correlates with peaks in HIV incidence in the sub-Saharan population. This suggests that the success of these lineages has been due in part to the increase in immune-compromised hosts in the areas of the invasive *Salmonella* Typhimurium epidemics (Okoro et al., 2012).

During 2011, an outbreak of bloody diarrhoea and haemolytic-uraemic syndrome (HUS) centred in Germany and southwest France was found to be caused by a clone of *E. coli* O104:H4 following acquisition of a Shiga-toxin encoding prophage and a plasmid encoding an extended spectrum β-lactamase. The acquisition of the Shiga-toxin gene by the O104:H4 lineage led to the emergence of a highly virulent strain causing significantly higher rates of HUS and mortality than most Shiga-toxin producing *E. coli* (Frank et al., 2011; Rasko et al., 2011; Rohde et al., 2011).

Human clinical practices can have a direct effect on the emergence of bacterial lineages. Phylogenetic analysis of 240 isolates representative of the globally disseminated *Streptococcus pneumoniae* PMEN1 lineage revealed high levels of recombination throughout the sampled population. Notably, 10 events were identified that resulted in a change of capsule type, with 1 such event resulting in a dramatic population shift coinciding with the introduction of a conjugate polysaccharide vaccine to the USA (Croucher et al., 2011). The vaccine escape serotype 19A has rapidly expanded, inhabiting the niche once occupied by vaccine-susceptible serotypes, demonstrating a rapid nationwide spread from the northeast of the United States (Golubchik et al., 2012). These data highlight the need to account for the range of diversity present within a bacterial population while designing public health interventions.

1.6. Summary

It is clear that WGS of populations of bacterial pathogens has the potential to vastly increase the knowledge of the evolutionary processes underpinning the dynamics of pathogen evolution on a scale ranging from infection of a single host, to the host population level. New insights gained from the study of pathogen evolution within a single host may aid decisions regarding appropriate clinical interventions. At the host population level, the ability to reconstruct transmission pathways on local and global scales can identify risk factors for the spread of pathogens, leading to the design or targeted public health programs aimed at minimising the impact of bacterial outbreaks.

1.7. Study aims

This work will evaluate the application of WGS to evolutionary and epidemiological studies of important bacterial pathogens. Specifically, the diversity among genome sequences from populations of *S. aureus* and *L. pneumophila* from individual hosts, *L. pneumophila* from a localised outbreak, and a pandemic clone of *S. aureus* will be examined in order to:

- Investigate the evolution of *S. aureus* within a single infected patient.

- Investigate the emergence and spread of the pandemic *S. aureus* CC30 lineage.

- Evaluate the use of whole genome sequencing for the resolution of a population of *L. pneumophila* isolates responsible for an epidemic of Legionnaire's disease.

2. Adaptive evolution of *S. aureus* during chronic endobronchial infection of a cystic fibrosis patient

2.1. Introduction

Cystic fibrosis (CF) is an autosomal recessive condition caused by mutations in the CF transmembrane conductance regulator (CFTR) protein. In healthy individuals the CFTR protein regulates transepithelial flow of ions, helping to maintain the proper ionic composition and volume of the airway surface fluid. The dysfunctional CFTR of CF patients leads to a change in the composition of the surface fluid which renders these patients susceptible to chronic endobronchial infections (Kerem et al., 1989; Rommens et al., 1989a, 1989b). At birth, the airways of CF patients are sterile but impaired mucociliary clearance, insufficient aeration of the paranasal sinuses and nasal polyps, along with frequent hospital visits mean CF patients are at particularly high risk of *S. aureus* infections (Gysin et al., 2000; Vu-Thien et al., 2010). Commonly, *S. aureus* is one of the earliest pathogens isolated from the sputum of CF patients, with a median time of first isolation of 214 days (range 120–309 days) (Souza et al., 2006). In CF patients, it is the oropharynx rather than the anterior nares that is the predominant site of *S. aureus* infection and persistence (Ridder-Schaphorn et al., 2007) and the upper airways have been identified as a reservoir allowing for recurrent infection of the lower airways by the same clone (Mainz et al., 2009). Further, longitudinal studies have shown that isolates belonging to the same lineage are repeatedly recovered from most CF patients, indicating chronic infection with a single strain (Branger et al., 1996; Vu-Thien et al., 2010). During chronic infection, *S. aureus* is subjected to numerous selective pressures resulting from antibiotic interventions, the host immune system and co-infection of the airways with other microorganisms (Harrison, 2007). Adaptive strategies of *S. aureus* are thought to involve the emergence of antibiotic-resistant, low-virulence, persistent phenotypes

(Goerke et al., 2010a; Kahl, 2010). Of note, a switch to small colony variants (SCV) is often observed in persistent infections and is associated with higher levels of resistance to antibiotics (Besier et al., 2007, 2008; Goerke et al., 2010a; Tuchscherr et al., 2011). In addition, hyper-mutable *S. aureus* strains with a defective DNA mismatch repair system have been isolated from CF infections which promote adaptation to encountered selective pressures (Oliver, 2010; Prunier et al., 2003, 2005). However, the molecular basis for the adaptation of *S. aureus* to the CF airways is not well understood and lags behind research into the CF pathogen *Pseudomonas aeruginosa* (Hogardt et al., 2010). For example, a comparative genomic study revealed evidence for selection acting across the whole *P. aeruginosa* genome, reflected in an adapted virulence phenotype with loss of function of factors required for acute infection but enhanced ability to cause chronic inflammation (Smith et al., 2006). Previous genome-scale analyses of *S. aureus* evolution during infection revealed that most mutations were the result of the strong selective pressure of antibacterial chemotherapy but may also be associated with promoting persistence (Gao et al., 2010; Mwangi et al., 2007).

2.2. Aims

- To understand the mechanism of *in vivo* adaptive diversification of *S. aureus* during chronic infection of the CF airways.
- To determine variation in phenotype, and attempt to link differences to observed polymorphisms.

2.3. Materials and methods

2.3.1. *S. aureus* strains

Sequential *S. aureus* strains were obtained from the sputum of a CF patient over a period of 26 months (Table 2.1). The strains had previously been characterised by PFGE, capsule- and phage-typing, and had been determined to belong to the same clone (Branger et al., 1996). No additional information was available regarding the patient, or antibiotic treatment administered during the course of infection.

2.3.2. *S. aureus* growth conditions

For isolation of individual colonies, culture was streaked on tryptic soy agar (TSA) plates, and incubated for 16 h at 37°C.

For growth to stationary phase, 5 ml tryptic soy broth (TSB) was inoculated with a single *S. aureus* colony and incubated for 16 h at 37°C with shaking at 200 rpm.

S. aureus cultures were stored in glycerol at -80°C.

2.3.3. *S. aureus* genomic DNA extraction

DNA was isolated from a 1 ml volume of stationary phase culture of *S. aureus* following the standard manufacturer's instructions using either the Bacterial Genomic DNA Purification Kit (Edge Biosystems, MD, USA), or the QIAcube Bacteria (Gram +) protocol, with the addition of 5µl 100µg/ml lysostaphin to the cell lysis step.

2.3.4. Growth in chemically defined media

To assess growth phenotype in nutrient depleted conditions, cultures grown in TSB overnight were used to inoculate chemically defined media (CDM) at a ratio of 200:1 and incubated at 37°C, with shaking at 200 rpm as described previously (Pohl et al., 2009). OD was measured at 30 m intervals using a FLUOstar OPTIMA plate reader (BMG Labtech, Germany).

CDM consisted of L-tryptophan (100 mg/l), L-tyrosine (100 mg/l), L- phenylalanine (100 mg/l), L-cysteine (50 mg/l), L-histidine (100mg/l), L-methionine (100 mg/l), L-glutamine (200 mg/l), L-glutamic acid (100 mg/l), glycine (100 mg/l), L-proline (100 mg/l), L- isoleucine (100 mg/l), L-leucine (100 mg/l), L-threonine (200 mg/l), L-valine (100 mg/l), DL-alanine (100 mg/l), L-arginine (100 mg/l), L-aspartic acid (100 mg/l), L-lysine (100 mg/l), hydroxy-L-proline (100 mg/l), L-serine (100 mg/l), ρ-aminobenzoic acid (0.2 mg/l), biotin (0.2 mg/l), folic acid (0.8 mg/l), niacinamide (1 mg/l), β-NAD (2.5 mg/l), pantothenate calcium salt (2 mg/l), pyridoxal (1 mg/l), pyridoxamine dihydrochloride (1 mg/l), riboflavin (2 mg/l), thiamine hydrochloride (1 mg/l), vitamin B12 (0.1 mg/l), adenine (20 mg/l), guanine hydrochloride (20 mg/l), uracil (20 mg/l), K_2HPO_4 (200 mg/l), KH_2PO_4 (1,000 mg/l), NaH_2PO_4 (3,195 mg/l), $MgSO_4$ (700 mg/l), $CaCl_2$ (10 mg/l), Na_2HPO_4 (9,214 mg/l) and glucose (10,000 mg/l). Where indicated, L-threonine was omitted for some experiments. The pH of the medium was buffered to 7.0.

2.3.5. Haemolytic activity assay

Haemolytic activity of strains was assessed after overnight culture of isolates at 37°C on TSA supplemented with sheep blood (Oxoid, UK).

2.3.6. Antibiotic sensitivity testing

Sensitivity to antibiotics including penicillin, oxacillin, gentamycin, kanamycin, tobramycin, ciprofloxacin, erythromycin, clindamycin, linezolid, teicoplanin, vancomycin, tetracycline, tigecycline, nitrofurantoin, fusidic acid, mupirocin, chloramphenicol, rifampicin, and trimethoprim, was performed using the Vitek 2 system (bioMérieux, Basingstoke, UK).

2.3.7. Genome sequencing

Genome sequencing was performed on the Illumina Genome Analyzer II platform. *S. aureus* strain ED83 was sequenced with 36 bp paired end reads, and strains ED84 and ED86 were sequenced with 36 bp single end reads. A mean coverage of 34.2 was achieved (range 25.7–44.1). Sequence reads for the 3 genomes have been deposited in the NCBI SRA database (accession number SRA038522).

Table 2.1 Origin and typing of strains used in this study

Isolate	Date of Isolation	MLST	Antibiogram	Reference
ED83	November 1983	30	Pn	Branger et al., 1996
ED84	October 1984	30	Pn	Branger et al., 1996
ED86	January 1986	30	PnFu	Branger et al., 1996

MLST, Multilocus Sequence Type; Pn, penicillin; Fu, fusidic acid

2.3.8. Sequence read alignment and identification of polymorphisms

Prior to alignment, 3 bp were trimmed from the 5' end of the sequence reads using the FASTX toolkit (Gordon, 2010), due to a GC % at these sites not consistent with the average GC % of the *S. aureus* genome.

Reads were aligned to the complete genome sequence of MRSA252 (accession number NC_002952) using Maq v0.7.1 (Li et al., 2008) and point mutations were identified from the pileup output from Maq at sites that varied from the reference sequence with a mapping quality greater than 30. Short insertions and deletions were identified using Novoalign (http://www.novocraft.com/main/index.php), with progressive trimming of unaligned single end reads by 2 bp to a minimum length of 15 bp. Polymorphisms were confirmed by Sanger sequencing of PCR products.

2.3.9. Identification of genetic content absent from the reference genome

Sequence reads that failed to map to the MRSA252 genome were assembled *de novo* using Velvet v0.7.63 (Zerbino et al., 2008), and protein-coding regions were identified through a BLASTx search against the NCBI non-redundant protein database with default settings.

Regions of difference were identified among the sequenced isolates using BRIG v0.71 (Alikhan et al., 2011) with MRSA252 as a reference, and the consensus sequence for each isolate as input.

2.3.10. Phylogenetic analysis

Maximum likelihood phylogenetic reconstruction was performed using RAxML v7.2.6 (Stamatakis, 2006) for all variable sites present in a nucleotide alignment of the study isolates and 5 representative CC30 strains (MRSA252 (Holden et al., 2004), WW2703_97, BTN1260, M1016 (Robinson et al., 2005a) and 07_1973 (McAdam et al., 2012)). A GTR model of nucleotide substitution was applied with the gamma model of rate heterogeneity. Support for nodes was assessed using 1000 bootstrap replicates.

2.4. Results

2.4.1. Sequential isolates are clonally related

In order to examine the microevolution of *S. aureus* during chronic endobronchial infection of a CF patient, the genomes of 3 sequential isolates obtained over 26 months from the airways of single patient were sequenced. In a previous study, the 3 isolates had been identified as being clonally related through a combination of PFGE, capsule- and phage-typing (Branger et al., 1996). Analysis of the MLST loci revealed each of the 3 isolates to belong to ST30, a globally distributed clone responsible for hospital and community associated epidemics over more than 6 decades (DeLeo et al., 2011; Robinson et al., 2005a).

Within the genomes of the 3 sequenced isolates a total of 23 point mutations and 15 insertion/deletions (indels) were identified, consisting of polymorphisms in 31 coding sequences and in 6 intergenic regions (Table 2.2). Phylogenetic analysis of the 3 sequential CF isolates in comparison to 5 representative strains of the same

clonal complex (CC30), based on core genome point mutations, indicates that the 3 CF isolates form a very narrow, distinct clade within the CC30 tree (Figure 2.1).

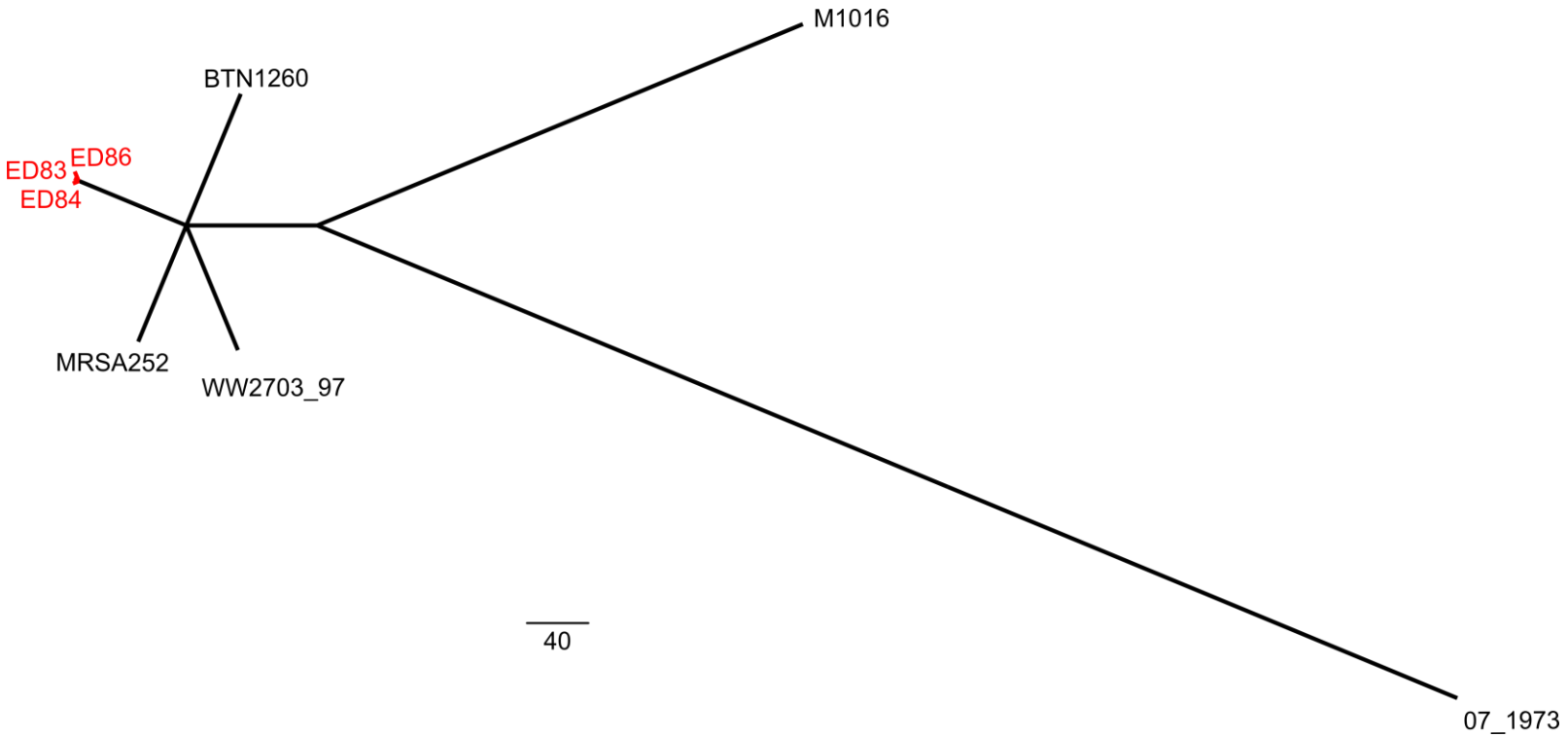


Figure 2.1 Unrooted maximum likelihood phylogeny based on core genome point mutations for the 3 CF isolates in comparison to 5 additional CC30 isolates. MRSA252 is representative of hospital associated epidemic ST36 strains (Holden et al., 2004), WW2703_97 is an ST30 MRSA strain, BTN1260 and M1016 are methicillin sensitive (MSSA) ST30 isolates from a global collection (Robinson et al., 2005a), 07_1973 is a PVL positive MRSA strain (McAdam et al., 2012). Scale bar represents a genetic distance of 40 polymorphisms.

2.4.2. The within-host *S. aureus* population is genetically heterogeneous

Of the 391 polymorphisms (376 point mutations and 15 indels) shared by the core genomes of all 3 isolates compared to the reference genome, none resulted in pseudogenes, and there were no mutations in the methyl-mismatch repair system (MutS and MutL) suggesting that the strains are not hypermutators.

Importantly, all 23 point mutations and 14 of the 15 indels identified among the 3 isolates were strain-specific, indicating clonal diversification of the original infecting strain leading to a genetically heterogeneous infecting population (Table 2.2).

The great majority of point mutations in coding sequences (94.7%) were non-synonymous, consistent with previous studies that have reported a high ratio of non-synonymous to synonymous point mutations within clonal complexes, in contrast to a higher frequency of synonymous point mutations between clonal complexes (Rocha et al., 2006).

2.4.3. Polymorphisms are found in loci associated with resistance, virulence and regulatory functions

Point mutations and indels were discovered in genes influencing global regulation, virulence, metabolism, and antibiotic resistance (Table 2.2).

In the last sequential isolate (ED86), 2 amino acid-altering point mutations were identified in the *fusA* gene encoding elongation factor G (EF-G) including a H457Y amino acid replacement previously demonstrated to confer fusidic acid resistance (Nagaev et al., 2001). Although the mutation is associated with a marked impairment in the biological fitness of *S. aureus*, the growth phenotype of ED86

compared favourably with that of the fusidic acid-susceptible isolates ED83 and ED84 in CDM (Figure 2.2). This may be due to a compensatory function of the second observed mutation, T187I, as has previously been reported in fusidic acid resistant strains of *S. aureus* (Nagaev et al., 2001). Consistent with the presence of these mutations, isolate ED86 was resistant to fusidic acid. All 3 isolates were resistant to penicillin but were sensitive to each of the other antibiotics tested.

Several of the identified mutations were in genes that may impact on virulence. For example, CodY is a global regulator and a potent repressor of virulence that links the metabolic state of the cell with virulence factor expression (Majerczyk et al., 2010; Pohl et al., 2009). The branched chain amino acid binding pocket of CodY is invariant in all finished *S. aureus* genome sequences available in the GOLD database (Figure 2.2). The gene encoding CodY in the third sequential isolate ED86 contains a non-synonymous point mutation resulting in a M65I amino acid replacement in the wall of the hydrophobic binding pocket which forms prominent hydrophobic interactions with the side chains of the branched chain amino acid ligand (Figure 2.3) (Levdikov et al., 2006). Of note, the ED86 isolate containing the M65I replacement has increased growth rate in CDM deficient in threonine in comparison to isolates ED83 and ED84 which both contain wild type CodY alleles (Figure 2.4), a growth characteristic previously demonstrated for a CodY *S. aureus* mutant (Pohl et al., 2009).

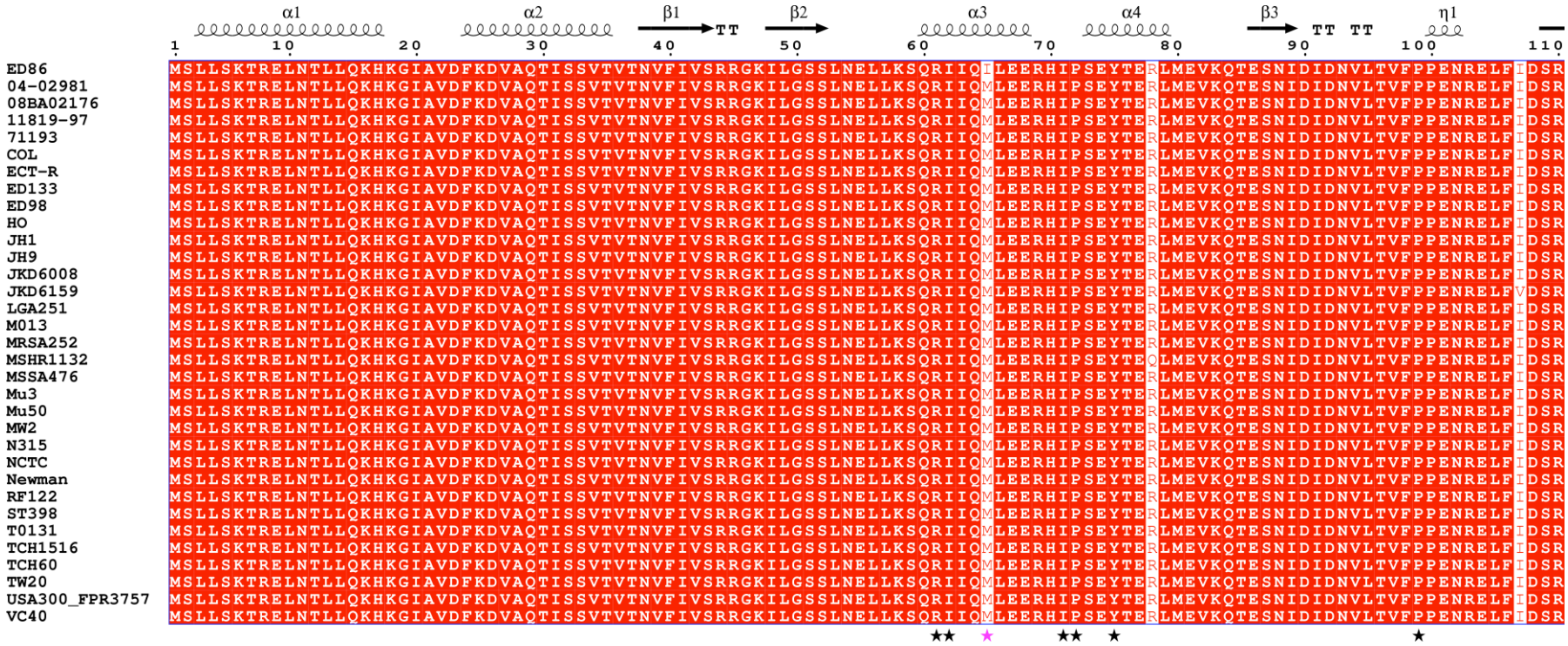


Figure 2.2 Alignment of the N-terminal domain amino acid sequence of CodY from 32 finished *S. aureus* genomes and the ED86 isolate. Invariant amino acids are highlighted in red. Stars indicate sites that form prominent interactions with branched chain amino acids in the hydrophobic binding pocket of CodY, the purple start indicates the variable residue in ED86. Secondary structure elements as identified in the crystal structure of CodY from *Bacillus subtilis* (PDB accession number 2B18) are indicated above the alignment. The alignment was generated using the program ClustalX2 (Larkin et al., 2007) and produced in ESPript (Gouet et al., 1999).

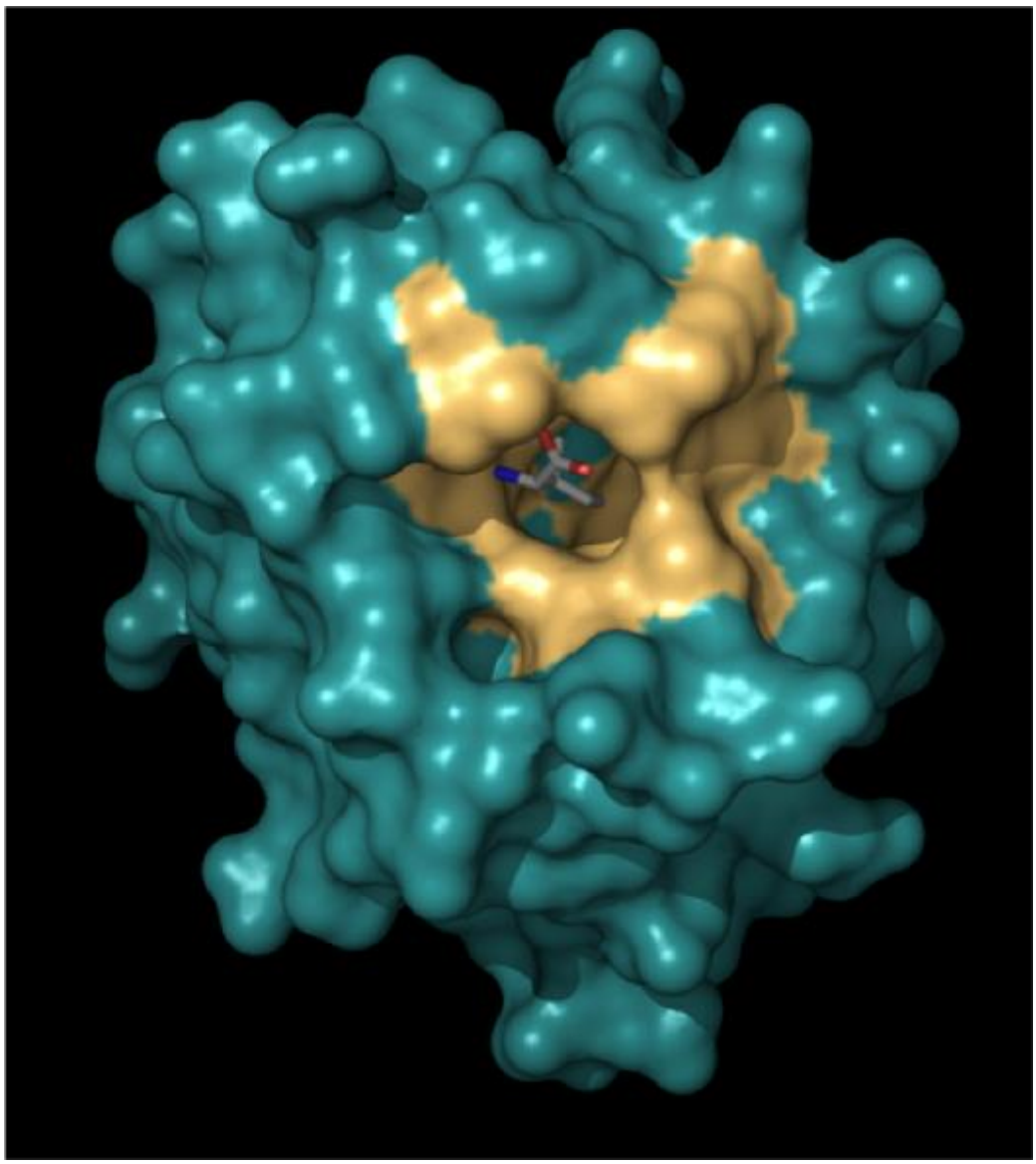


Figure 2.3 3D surface representation of the branched chain amino acid binding pocket of CodY (Levdikov et al., 2006). The non-synonymous substitution observed in isolate ED86 results in an amino acid change within the wall of the binding pocket, highlighted in orange.

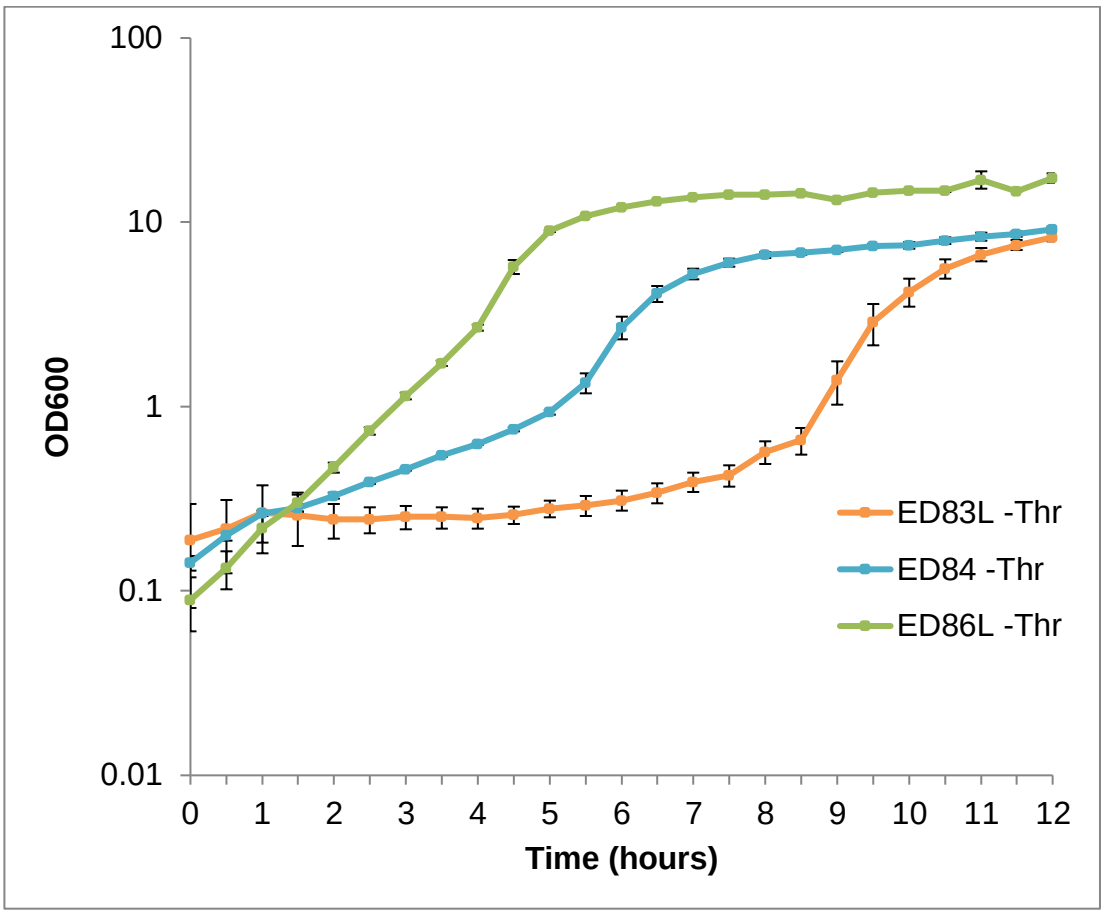


Figure 2.4 Growth curve of *S. aureus* isolates ED83 (large colony), ED84, and ED86 (large colony) in chemically defined media lacking threonine at 37°C

2.4.4. Multiple independent polymorphisms in the SigB regulon suggest these loci are subject to strong selective pressures in this patient

Multiple, independent polymorphisms were observed in all 3 isolates which would influence expression of SigB, a regulator of the general stress response that influences the expression of virulence factors. Specifically, a frameshift mutation resulting in a *sigB* pseudogene in strain ED83, a frameshift mutation in SpoVG, a proposed modulator of SigB activity (Meier et al., 2007), and an 18 bp in-frame deletion in the *rsbU* regulator of SigB in ED84, were identified. Comparison of the ED84 *rsbU* allele with the crystal structure of the *Bacillus subtilis* RsbU protein (PDB accession number 2J70) suggests that the 18 bp deletion is likely to impair the phosphatase activity of RsbU.

In addition, SpoVG and the histidine kinase ArlS, which have been shown to impact SigB-dependent capsule formation (Meier et al., 2007), are affected by a non-synonymous point mutation and a frameshift insertion respectively in the ED86 isolate. The parallel evolution of mutations at SigB-associated loci in the 3 isolates represent a 119-fold increase in mutation rate when compared to the mutation rate across the whole genome ($p < 0.05$). In order to further investigate this phenomenon, the DNA sequence for SigB-associated loci *rsbU*, *rsbV*, *rsbW*, *sigB*, *spoVG*, *arlS*, and *arlR* was determined for a panel of 16 *S. aureus* isolates from 4 CF patients persistently colonized with *S. aureus* (Branger et al., 1996). However, only 2 mutations (non-synonymous substitutions in the *spoVG* and *arlR* loci in different isolates) were observed, and no pseudogene forming mutations were identified at any of the loci examined.

2.4.5. Phenotypic variation was observed in colony size, haemolytic activity, and growth rate in minimal media

Subculturing of single colonies of isolates ED83 and ED86 onto TSA containing sheep erythrocytes resulted in 2 distinct colony sizes that differed in haemolytic activity (large colonies were haemolytic and small colonies were non-haemolytic) revealing an unstable *in vitro* phenotype (Figure 2.5). Subculturing of ED84 yielded only one colony morphology (haemolytic) (Figure 2.5).

A difference in the growth rates in CDM deficient in threonine was observed for large and small variants of ED83 and ED86, with the large variant of each isolate having at a faster growth rate than the small variant (Figure 2.6).

Of note, fitness as assessed by growth in CDM lacking threonine has increased in isolates from later time points during infection (Figure 2.4).

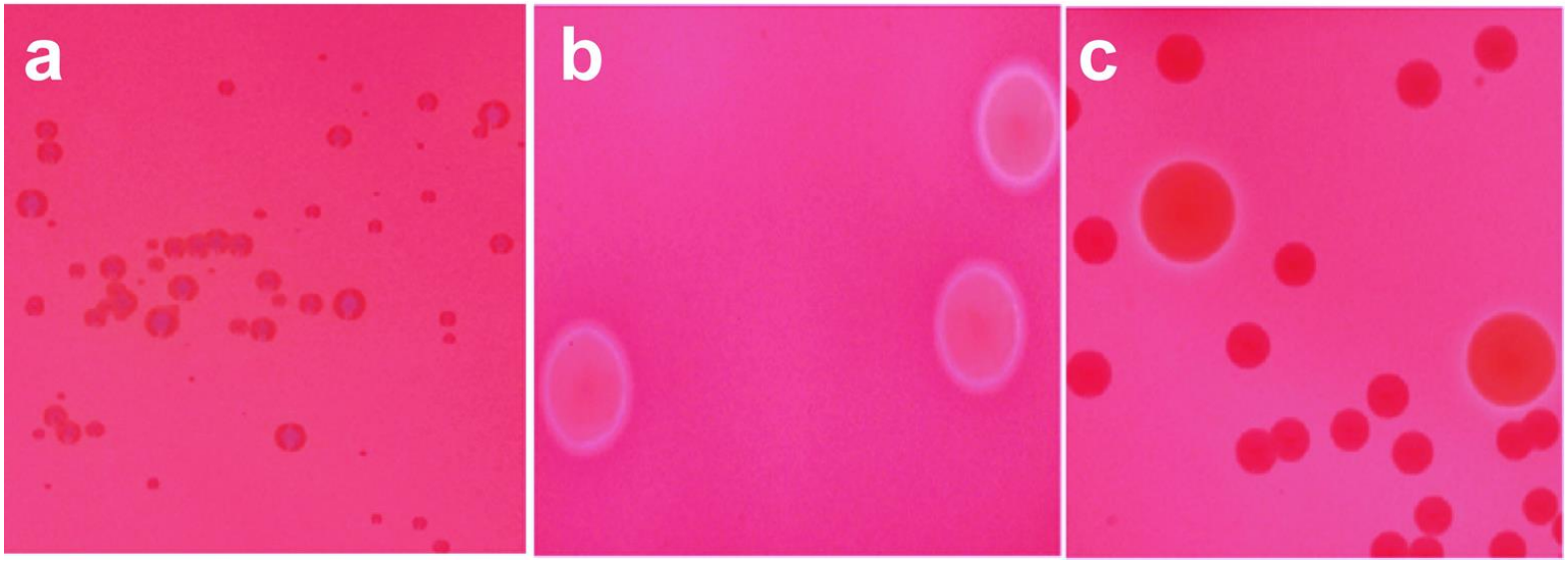


Figure 2.5 Haemolytic activity on tryptic soy agar supplemented with sheep erythrocytes for (a) colony variants of ED83, (b) ED84 and (c) colony variants of ED86

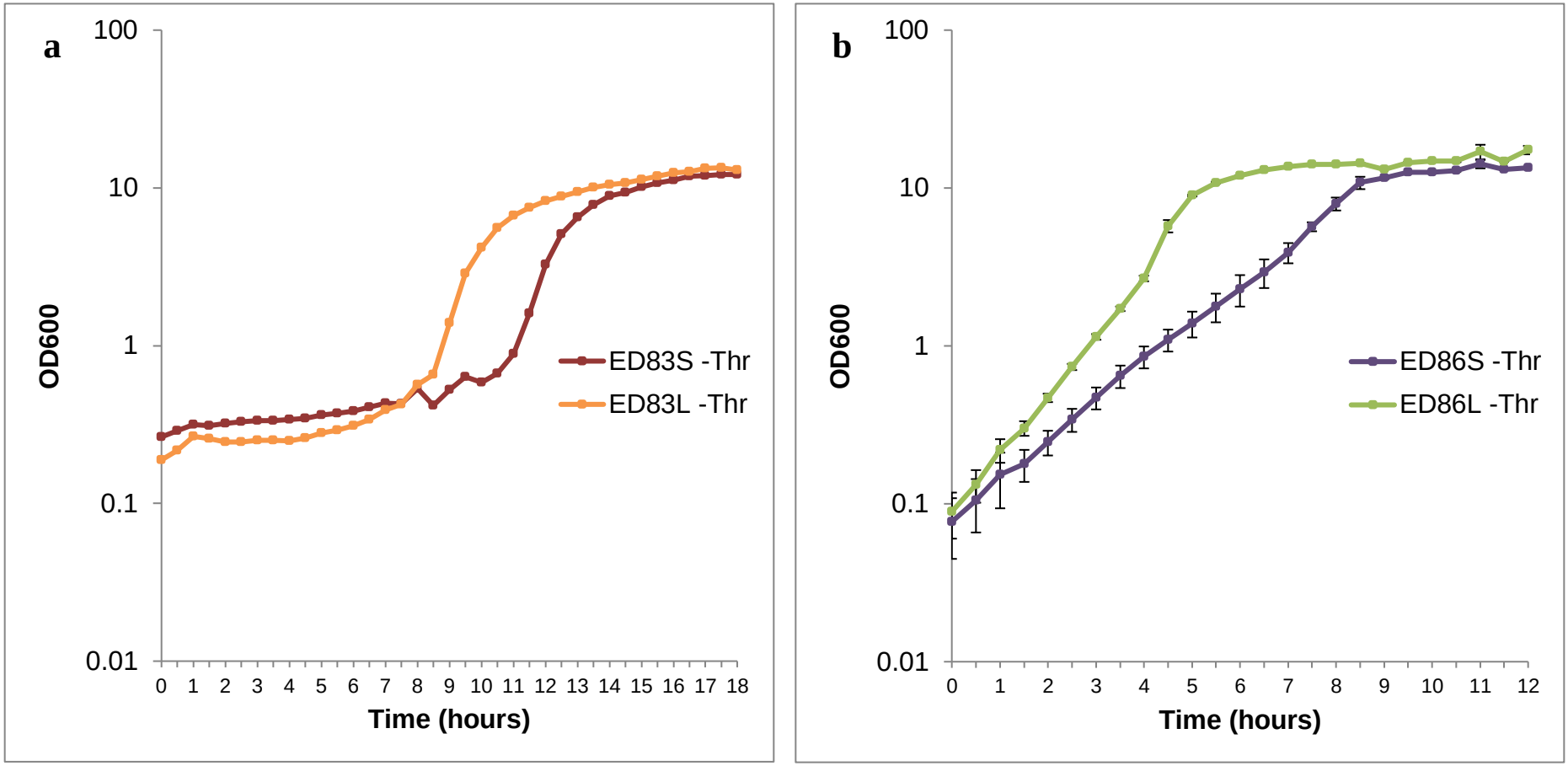


Figure 2.6 Growth curve of large (L) and small (S) variants of (a) ED83 and (b) ED84 in chemically defined media lacking threonine at 37°C

2.4.6. Phage content varies among the study isolates

Examination of the genome sequence of isolate ED83 revealed 3 prophages including a 48.3 kb phage related to ϕSa2 of MRSA252, a 43.5 kb β-converting phage encoding the secreted virulence factors staphylococcus enterotoxin A (SEA) and staphylokinase (SAK), and a 20.2 kb phage remnant with homology to ϕNM4 of *S. aureus* strain Newman. Comparative genomic analysis indicates that isolate ED84 had lost the β-converting phage whereas isolate ED86 lacked the ϕSa2 homologue (Figure 2.7). Consistent with absence of the β-converting phage, strain ED84 demonstrated haemolytic activity for sheep erythrocytes, indicating the presence of a functional β-toxin gene (Figure 2.5).

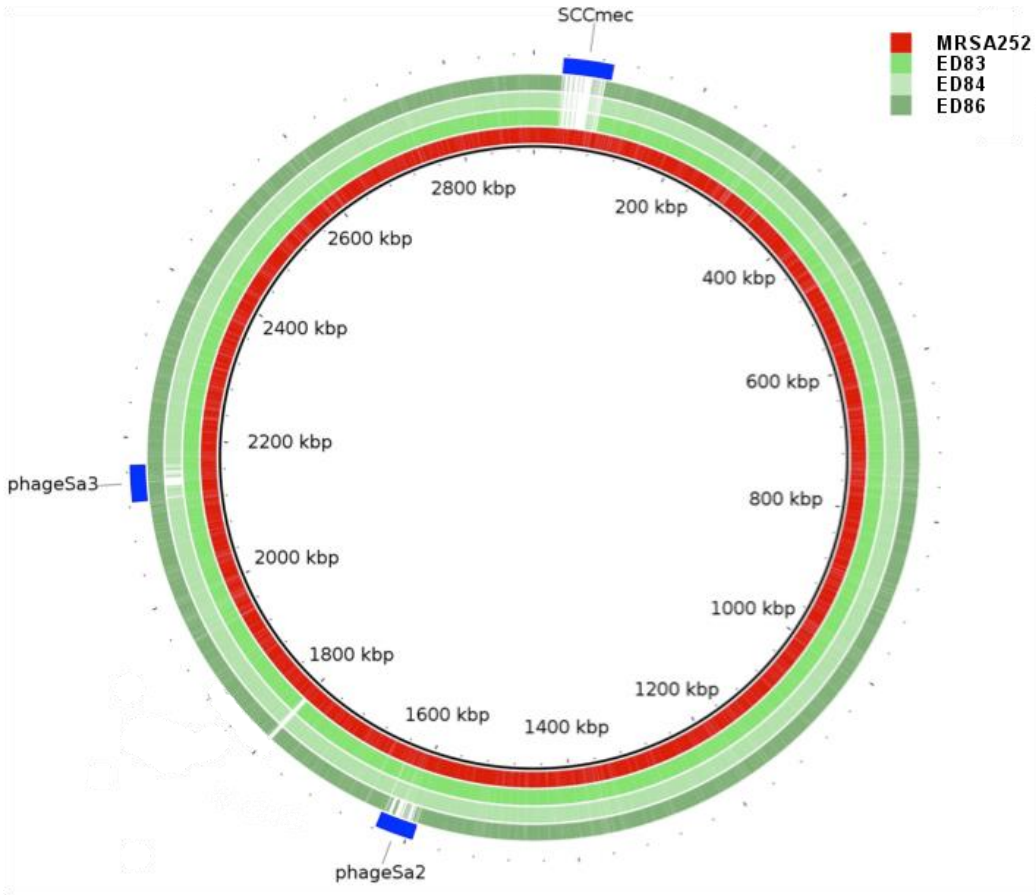


Figure 2.7 Schematic genome diagram of ED83, ED84, and ED86 with reference to MRSA252. Gaps in alignment correspond to mobile genetic elements that are annotated in the outermost ring.

Table 2.2 Genetic polymorphisms identified among study isolates

Isolate	Polymorphism	Polymorphism effect[1]	Locus
ED83	Point mutation	NSS	Hypothetical protein
ED83	Point mutation	NSS	*yycG* (Sensor kinase protein)
ED83	Point mutation	NSS	Putative cobalmin synthesis protein
ED83	Point mutation	NSS	Putative acetyl transferase
ED83	2 point mutations in single codon	NSS	*sucA* (2-oxoglutarate dehydrogenase E1 subunit)
ED83	1bp deletion	Frameshift affecting 46.6% of protein	Hypothetical membrane protein
ED83	1bp insertion	Frameshift affecting 18.5% of protein	*rpsI* (30S ribosomal protein S9)
ED83	1bp insertion	Frameshift affecting 80.8% of protein	Hypothetical protein
ED83, ED86	1bp deletion	-	Intergenic region
ED83	1bp deletion	Frameshift affecting 13.4% of protein	Hypothetical protein
ED83	1bp deletion	Frameshift affecting 23.1% of protein	*phoR* (alkaline phosphatase synthesis sensor)
ED83	1bp insertion	Frameshift affecting 36.7% of protein	*sigB* (RNA polymerase sigma factor SigB)
ED84	Point mutation	-	Intergenic region

Isolate	Polymorphism	Polymorphism effect[1]	Locus
ED84	Point mutation	NSS	ABC transporter ATP-binding protein
ED84	Point mutation	NSS	*glmM* (putative phsphoglucosamine mutase)
ED84	Point mutation	-	Intergenic region
ED84	Point mutation	NSS	*fnbA* (Fibronectin-binding protein precursor A)
ED84	Point mutation	-	Intergenic region
ED84	1bp deletion	Frameshift affecting 89.0% of protein	*spoVG* (regulatory protein SpoVG)
ED84	18bp deletion	Deletion of 6 amino acids	*rsbU* (putative sigB regulation protein)
ED84	1bp insertion	Frameshift affecting 85.4% of protein	Hypothetical protein
ED84	1bp insertion	-	Intergenic region
ED84	1bp insertion	Frameshift affecting 46.3% of protein	Putative membrane protein
ED86	Point mutation	NSS	*spoVG* (Regulatory protein SpoVG)
ED86	Point mutation	NSS	*fusA* (Elongation factor G)
ED86	Point mutation	NSS	*fusA* (Elongation factor G)
ED86	Point mutation	SS	Hypothetical membrane protein

Isolate	Polymorphism	Polymorphism effect[1]	Locus
ED86	Point mutation	NSS	*codY* (Transcriptional repressor CodY)
ED86	Point mutation	NSS	*infB* (Translation initiation factor IF-2)
ED86	Point mutation	NSS	Prephenate dehydrogenase
ED86	Point mutation	NSS	Putative peptidase
ED86	Point mutation	NSS	*ald1* (Alanine dehydrogenase)
ED86	Point mutation	NSS	*moaB* (Putative molybdenum cofactor biosynthesis protein B)
ED86	Point mutation	NSS	Transcriptional regulator (antiterminator)
ED86	56bp deletion	-	Intergenic region
ED86	1bp deletion	Frameshift affecting 97.4% of protein	Hypothetical membrane protein
ED86	1bp insertion	Frameshift affecting 41.2% of protein	*arlS* (Sensor kinase protein)

[1]NSS; Non-synonymous substitution, SS; Synonymous substitution

2.5. Discussion

With the ability to rapidly sequence bacterial pathogen populations, it is possible to carry out real-time studies of pathogen evolution during long-term infection. Genome sequencing of 3 sequential isolates from a single CF patient showed the isolates were clonally related, in agreement with results from PFGE, capsule- and phage typing methods (Branger et al., 1996). However, the resolution afforded by whole genome polymorphism identification reveals that the level of diversity present within the infecting *S. aureus* population is consistent with evolution following a single transmission event (Golubchik et al., 2013).

Pathogenic bacteria have a close association with their host, being subject to selective pressures resulting from co-infection with other microorganisms, the immune system of the host, and antibiotic therapies (Harrison, 2007). Adaptive strategies of bacteria to long-term colonisation are thought to include the emergence of antibiotic resistant, low virulence phenotypes that are able to establish chronic infection, without causing mortality in the host (Gao et al., 2010; Mwangi et al., 2007; Smith et al., 2006). Selection for antibiotic resistant phenotypes, and the molecular mechanism behind the resistance, have been demonstrated in *S. aureus* through the use of whole genome sequencing of isolates obtained from a patient undergoing antibiotic therapy for infective endocarditis (Mwangi et al., 2007). A comparative genomic analysis of an antibiotic sensitive isolate taken before the initiation of antimicrobial therapy with a resistant isolate obtained 3 months later after treatment with rifampin, imipenem, and vancomycin revealed a total of 35 point

mutations in 31 loci. At least 6 loci had previously been implicated in conferring resistance to a range of antibiotics.

The identification of a fusidic acid resistant phenotype in isolate ED86, and that this strain harbours an allele of *fusA* previously shown to confer resistance to fusidic acid (Nagaev et al., 2001) highlights the potential of sequencing technologies in the clinical setting. It is now possible to obtain genome bacterial genome sequences in a clinically relevant timeframe (Köser et al., 2012), and it has been demonstrated that the antibiotic resistance profile of an important clone of *S. aureus* can be identified from the genome sequence alone, with an accuracy of 99.8% (Holden et al., 2013).

The high percentage of non-synonymous polymorphisms identified in coding regions (94.7%) supports the hypothesis that mildly deleterious mutations are maintained in the short term. However, the high proportion of non-synonymous mutations are likely, at least in part, to be the result of strong selective pressures that exist in the unique environment of the CF airways. Evolution from pathogenicity to commensalism is beneficial for the long-term survival of a microbe during the course of chronic infection of the host. Within the 3 study isolates, a number of polymorphisms are identified in regulators of virulence.

Multiple, independent polymorphisms were found in the SigB regulon. SigB is a regulator of the general stress response that influences the expression of virulence factors (Meier et al., 2007). The level of pseudogene formation within this regulon suggests that a strong selective pressure is acting on these loci leading to beneficial phenotypes adapted to the environment of the CF airways.

Several previous studies have identified a role for SigB in the development of SCVs and in the regulation of transcription of virulence factors during infection of CF patients (Goerke et al., 2010b; Meier et al., 2007; Moisan et al., 2006). Data from this study suggest that there may be selective pressure for attenuation or loss of SigB expression during long-term infection. However, the screening of isolates from an additional 4 patients with chronic *S. aureus* infection did not identify any evidence for selection within the SigB regulon. Taken together, these data suggest that the elevated mutation rate in SigB-associated loci in *S. aureus* isolates ED83, ED84 and ED86 is related to the genetic background of the strain or to patient-specific selective pressures, and is not broadly applicable to all CF *S. aureus* isolates. Of note, clinical isolates of *S. aureus* with SigB deficiency have been reported previously (Karlsson-Kanth et al., 2006). Entenza and colleagues investigated the role of SigB in a rat model of infective endocarditis and discovered increased bacterial densities of SigB-deficient strains after prolonged infective endocarditis infection suggestive of a selective pressure for reduced SigB activity during infection (Entenza et al., 2005).

In addition to polymorphisms of the core genome, variation in the accessory genome of the isolates was observed. Phage content was variable among the 3 isolates (Figure 2.7), indicating that phage deletion and/or acquisition events occur during *S. aureus* infection of the CF airways, consistent with previous findings (Goerke et al., 2004). By maintaining phages that encode secreted virulence factors in a subpopulation of the infecting bacteria, the metabolic costs to the infecting population as a whole are reduced in nutritionally-starved niches.

It remains to be determined whether the observed genetic changes in observed in isolates from CF patients result in changes in fitness and transmissibility outside of the host, but recent work has shown that genetic variants of *P. aeroginosa* that emerge during growth in persistent biofilms have impaired fitness outside of this niche (Penterman et al., 2014).

2.6. Conclusions

The whole genome approach to the study of chronic bacterial infection offers a high-resolution view of the process of within-host adaptive evolution. Understanding the molecular basis for changes in antibiotic resistance profiles and in virulence factor expression offers a platform for the development of rapid diagnostic assays and personalised therapies.

Overall, the genome-scale identification of polymorphisms affecting antibiotic resistance, growth, and global regulation of virulence indicate a profound impact on the ecology of *S. aureus* during chronic infection. The identification of loci under selective pressure during chronic *S. aureus* infection may indicate novel therapeutic targets for the control of persistent infections that are refractory to treatment.

3. Evolution and epidemiology of a pandemic clonal lineage of *S. aureus*

3.1. Introduction

Staphylococcus aureus is a component of the normal flora of about 30% of the human population, but is capable of causing severe infections of immunocompromised patients in hospitals and healthy humans in the community (Wertheim et al., 2005). Hospital-associated methicillin-resistant *S. aureus* (HA-MRSA) is represented by a small number of clones that rarely cause disease outside of the healthcare setting and that are characterised by resistance to β-lactam antibiotics, in addition to other front-line antimicrobials (Chambers et al., 2009). During the last 6 decades, the *S. aureus* clonal complex 30 (CC30) has had a profound impact on global human health by giving rise to three pandemic waves and the toxic shock syndrome (TSS) epidemic (Altemeier et al., 1982; Robinson et al., 2005a). Furthermore, one study indicated *S. aureus* isolates from life-threatening endocarditis infections are more likely to belong to CC30 than to other *S. aureus* lineages (Nienaber et al., 2011). The first CC30 pandemic was caused by methicillin-sensitive phage type 80/81 clone in the 1950s and 1960s, which spread from hospitals causing a significant disease burden in the community and was characterized by resistance to penicillin and production of the PVL toxin (Donahue et al., 1966; Goldie et al., 1971; Nahmias et al., 1961). The Southwest Pacific clone (SWP) is a contemporary PVL positive community-associated MRSA clone, which has spread to several continents and which largely causes skin and soft tissue infections of otherwise healthy individuals (Alesana-Slater et al., 2011; Collignon et al., 1998). In contrast to phage type 80/ 81 and SWP, the EMRSA-16 (ST36) clone appears to be restricted to the hospital setting and has reduced virulence due in part to low levels of expression of cytolytic toxins (DeLeo et al., 2011). Along with the

EMRSA-15 (ST22) clone, EMRSA-16 has been endemic in UK hospitals for over 20 years and has also been reported less commonly in other European countries, Southeast Asia, South Africa, Australia, and North America (Carleton et al., 2004; Ho et al., 2009; Jansen van Rensburg et al., 2011; Robinson et al., 2005a; Scicluna et al., 2010). The high rate of MRSA infections and the rapid spread of HA-MRSA between UK hospitals led the UK government to introduce stringent infection control legislation from 2003, resulting in a decrease in rates of nosocomial MRSA infection. Of note, EMRSA-16 prevalence has declined more rapidly than that of EMRSA-15, implying the existence of unknown strain-dependent factors that confer increased susceptibility to hospital infection prevention and control measures (Ellington et al., 2010; Wyllie et al., 2011). Despite its clinical importance, the evolution of the EMRSA-16 clone, in addition to the molecular basis for its success are poorly understood. A phylogenomic approach was used to examine the diversity of EMRSA-16 relative to other CC30 pandemic clones. The results represent a high-resolution insight into the emergence, expansion, and transmission of a major clone of MRSA, revealing unique molecular correlates for its hospital association.

TSS is a severe toxinoses caused largely by *S. aureus* strains that produce the superantigen toxic shock syndrome toxin-1 (TSST-1) (Bergdoll et al., 1981). The TSST-1 gene (*tst*) is encoded on the Staphylococcal Pathogenicity Island (SaPI2), which is variably present in the *S. aureus* population (Subedi et al., 2007). A major epidemic of TSS caused largely by strains of *S. aureus* CC30 occurred in the 1970s-80s in menstruating women using a particular brand of hyper-absorbent tampons (Fitzgerald et al., 2001; Shands et al., 1980), although TSS can occur in non-menstruating women, and in males (Davis et al., 1980). Although the incidence of

TSS in the population has fallen significantly since the 1980s, the disease still poses a significant clinical problem (DeVries et al., 2011; Osterholm et al., 1982).

3.2. Aims

- Determine the evolutionary history of the CC30 lineage.
- Identify molecular correlates of CC30 pandemic lineage emergence and adaptation.
- Improve the understanding of the diversity and transmission routes of EMRSA-16 relative to other CC30 clones.
- Determine the distribution of TSST-1 within the CC30 lineage, and identify those clones responsible for causing TSS.

3.3. Materials and methods

3.3.1. Strains selected for this study

In addition to 7 publicly available complete or draft CC30 genome sequences, a total of 80 *S. aureus* isolates were selected from a larger collection for genome sequencing to represent the breadth of genotypic diversity within the CC30 lineage sampled through time and space. The genetic diversity of CC30 isolates was based on molecular typing methods such as pulsed field gel electrophoresis (PFGE), staphylococcal protein A (*spA*), and MLST, in addition to limited information relating to the distribution of selected virulence and antibiotic resistance determinants. Strain collection was focused primarily on EMRSA-16 isolates from the United Kingdom. The completed strain collection consisted of 87 isolates in total, including 30 isolates from England, 29 from Scotland, 12 from USA, 5 from

Australia, 3 from France, 2 from Denmark and Finland, and 1 each from Germany, Hungary, Republic of Ireland and Sweden. The 80 isolates sequenced for the current study were typed by MLST as ST30 (n = 21) or its single locus variant (slv) ST36 (n = 58), with one isolate identified as ST500, a single locus variant of both ST30 and ST36.

3.3.2. *S. aureus* growth conditions

For isolation of individual colonies, culture was streaked on tryptic soy agar (TSA) plates, and incubated for 16 h at 37°C.

For growth to stationary phase, 5 ml tryptic soy broth (TSB) was inoculated with a single *S. aureus* colony and incubated for 16 h at 37°C with shaking at 200 rpm.

S. aureus cultures were stored in glycerol at -80°C.

3.3.3. *S. aureus* genomic DNA extraction

DNA was isolated from a 1 ml volume of stationary phase culture of *S. aureus* following the standard manufacturer's instructions using either the Bacterial Genomic DNA Purification Kit (Edge Biosystems, MD, USA), or the QIAcube Bacteria (Gram +) protocol, with the addition of 5µl 100µg/ml lysostaphin to the cell lysis step.

3.3.4. Whole genome sequencing

Whole genome sequencing of bacterial isolates was performed by ARK-Genomics (The Roslin Institute, University of Edinburgh). Sequencing libraries were prepared using the standard TruSeq or Nextera protocols for multiplexed single or paired end

reads (Illumina, San Diego, CA). Short DNA reads were produced using the Illumina GA2 platform.

3.3.5. Pre-assembly read processing

Raw FASTQ files were examined for Illumina adaptor sequences using cutadapt v1.2 (Martin, 2011) to facilitate removal of adaptor contamination. The 5′ and 3′ ends of reads were trimmed to remove low quality scoring bases (Q <30) using Sickle v1.2 (https://github.com/najoshi/sickle). Reads containing low frequency sequencing errors were corrected with Quake v0.3 using a kmer size of 15 (Kelley et al., 2010).

3.3.6. Mapping of sequence reads

Reads from the previously sequenced ED83, ED84, ED86 (McAdam et al., 2011) and WW2703/97 (Thomas et al., 2011) isolates were included in this study. Adaptor sequences were trimmed from Illumina reads using the ea-utils FASTQ processing tool (Aronesty, 2011). Illumina reads were mapped to the ST36 MRSA252 genome sequence (Holden et al., 2004) using the Burrows–Wheeler short-read aligner (BWA) with the Smith–Waterman alignment of unmapped mates disabled for paired end reads (Li et al., 2009). Reads generated on the 454 platform were trimmed using the Biopython SeqIO module and mapped to MRSA252 with the BWA long-read aligner (Li et al., 2010). Consensus sequences were called and point mutations and insertions/deletions (indels) identified for sites covered by at least 3 reads, with average mapping and PHRED scores greater than 30. Consensus genomes and whole genomes representative of the EMRSA-16 clone (MRSA252), SWP clone (TCH60, acc. no. CP002110.1), and other epidemic CC30 (MN8, accession number

CM000952.1) were aligned using the progressiveMauve algorithm and gap positions removed (Darling et al., 2010). Sites in the core genome were defined as those sites present in all sequenced isolates.

The presence of the PVL locus was assessed by mapping sequence reads to previously sequenced phage with either an elongated (ΦSa2958, accession number AP009363) or icosahedral (Φ108PVL, accession number AB243556) head morphology. The presence of the SapI2 integrase and TSST-1 was assessed by mapping sequence reads to the SaPI2 sequence from strain RN3984 (accession number EF010993).

3.3.7. Bayesian phylogenetic analysis

Bayesian analysis of evolutionary rates and divergence times was performed using BEAST v1.6.1 (Drummond et al., 2007) with the Hasegawa–Kishino–Yano model of nucleotide substitution and a gamma correction for rate heterogeneity. The SWP and Phage 80/81 clades were constrained together on the basis of a robust, maximum likelihood phylogeny rooted using an ST5 outgroup. All isolates were dated on the basis of year and month (when known) of isolation. Before carrying out the Bayesian analysis, the dataset was assessed for temporal signal using a linear regression of root-to-tip branch length against date of isolation for the maximum likelihood phylogeny (Figure 3.1). Markov chain Monte Carlo (MCMC) samples from three independent analyses each run for 1.5×10^8 iterations, sampled every 1,000 generations, with the first 10% discarded as burn-in, were combined for estimation of posterior probabilities. The lognormal relaxed molecular clock model was used, as a Bayes Factor test favoured this over a strict clock, with a constant

coalescent prior (Drummond et al., 2006). The presence of recombinant regions in the alignment was assessed using the phi test in the SplitsTree package (Huson et al., 2006), and the consistency index for the phylogeny was calculated in PAUP (Wilgenbusch et al., 2003).

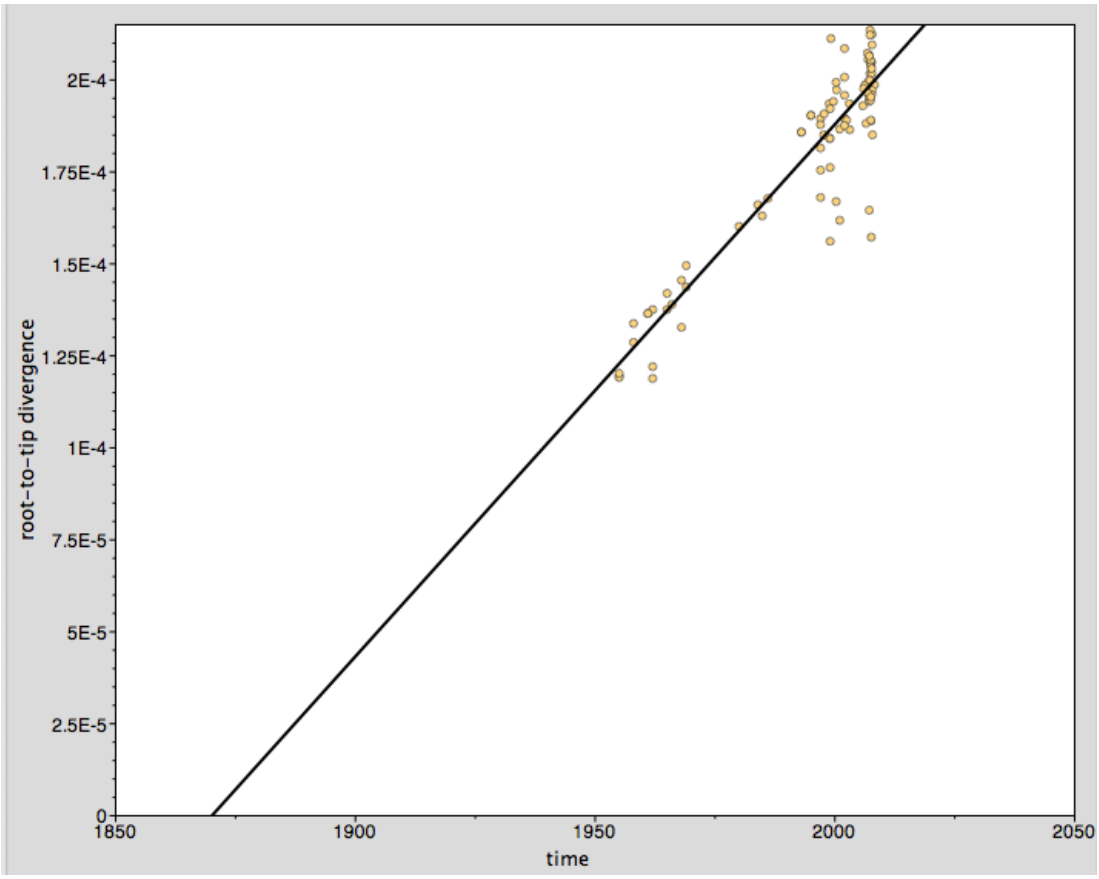


Figure 3.1 Linear regression of root-to-tip branch length against date of isolation for the maximum likelihood phylogeny of CC30 was performed using Path-O-Gen v1.3 (http://tree.bio.ed.ac.uk/software/pathogen/). Gradient is equivalent to rate of molecular evolution. Gradient = $1.46E^{-6}$, $R^2 = 0.86$

3.3.8. Bayesian phylogeographic analysis

For an alignment of the UK isolates of the EMRSA-16 clone, phylogeographic distribution was examined using the discrete diffusion model (Lemey et al., 2009) with distance-informed priors and using the clock rate determined for the whole dataset. Isolates were grouped by geographic region (London, southeast, south, and central England and north, east, and west Scotland). A Bayes factor of 3.0 was set as the limit for significant links between locations.

3.4. Results

3.4.1. Phylogenetic analysis of CC30 strains reveals 3 distinct pandemic clades and a fourth associated with community-acquired infections

The core genome of the 87 CC30 isolates examined consisted of 2,381,276 bases (82% of the MRSA252 reference genome), containing a total of 4,499 high-confidence polymorphic sites. The application of a Bayesian coalescent method using a relaxed molecular clock model to infer the phylogeny of the CC30 lineage indicated the existence of 3 major clades within the CC30 lineage, representative of the major pandemic clones, 80/81, SWP, and EMRSA-16, in addition to an EMRSA-16 sister clade represented by other epidemic CC30 isolates (Figure 3.2). There was strong posterior support for the majority of nodes in the tree and parsimony analysis indicated a low frequency of homoplasies across the phylogenetic tree (consistency index of 0.92 for parsimony informative sites). The phi test for recombination

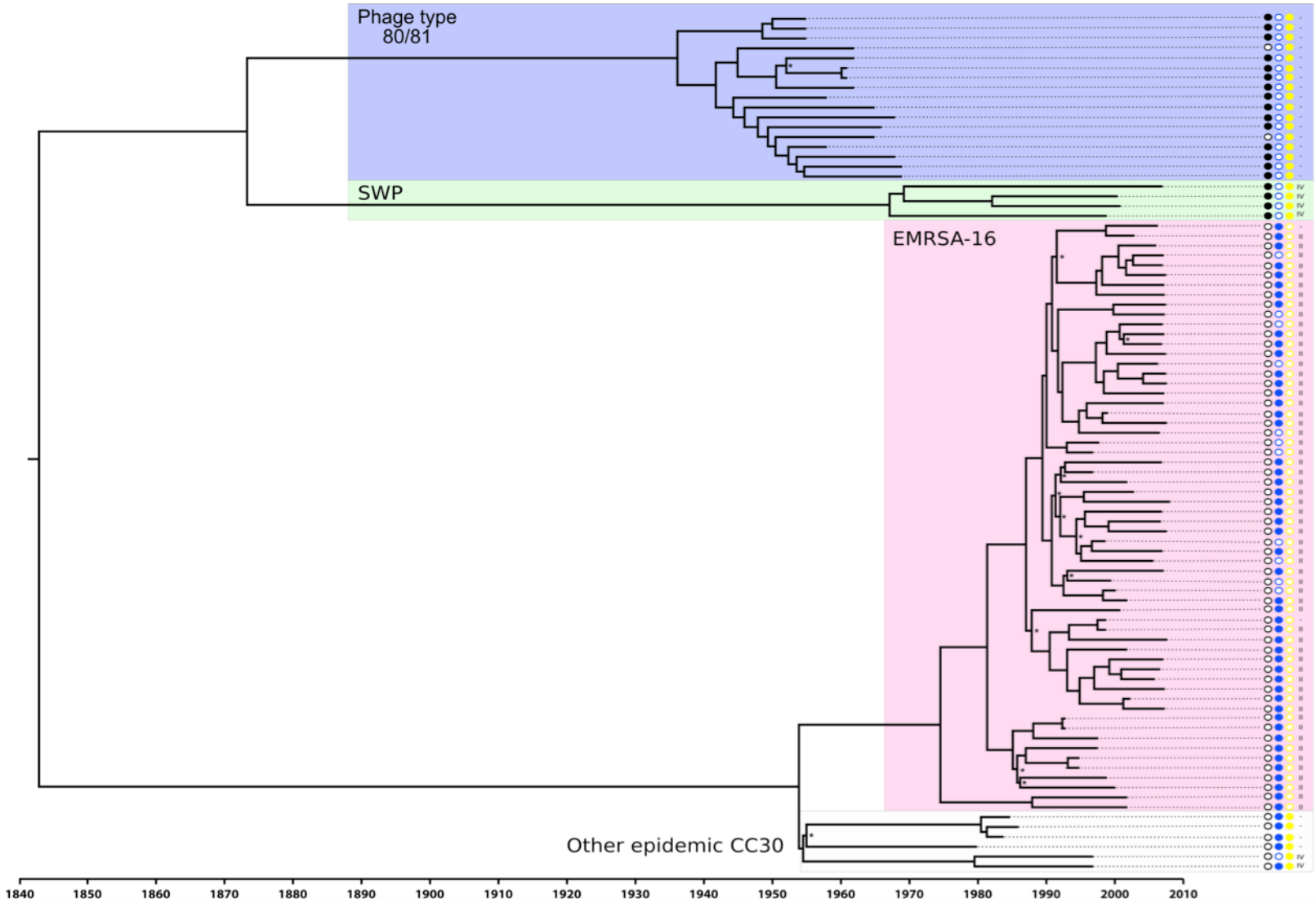

Phage type
80/81
SWP
EMRSA-16
Other epidemic CC30
1840
1850
1860
1870
1880
1890
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010

Figure 3.2 Bayesian phylogenetic reconstruction of the CC30 lineage using all sites in core genome. The lineage is divided into multiple distinct clades, characterised by the presence or absence of different virulence and antimicrobial resistance determinants. Blue, green, red, and grey shading correspond to the 80/81, southwest Pacific, EMRSA-16, and other epidemic CC30 clones respectively. Presence of the *pvl* loci is denoted by shaded black circles, *tst* carriage by shaded blue circles, intact *crtM* gene by shaded yellow circles, and SCC*mec* type is indicated for MRSA isolates. Branch lengths are scaled with respect to time. Nodes have posterior probabilities of >0.8 unless indicated with an asterisk. Tip and node labels are shown in appendix 2.

implemented in SplitsTree returned a p-value > 0.05, indicating the absence of significant levels of recombination in the dataset.

3.4.2. Bayesian evolutionary analysis determines the rate of the molecular clock for CC30 and facilitates estimates of the times of emergence of pandemic clones

The rate of the molecular clock was estimated simultaneously with the phylogeny. The mean nucleotide substitution rate within CC30 was 1.42×10^{-6} substitutions per site per year [95% highest posterior densities (HPDs) 1.04×10^{-6} to 1.80×10^{-6}]. Given the rate of molecular evolution determined for the CC30 lineage, the calculated time of the most recent common ancestor (MRCA) for each of the 3 major clades that correspond to the 3 pandemic CC30 clones was calculated. The date for the MRCA of the 80/81 clone was estimated as 1936 (95% HPDs, 1926–1945), the date for the MRCA of the SWP clone was estimated as 1967 (95% HPDs, 1952–1984), and the date for the MRCA of the EMRSA-16 clone was estimated as 1975 (95% HPDs, 1965–1983). Finally, the date for the MRCA of the entire CC30 lineage was estimated as 1842 (95% HPDs, 1765–1909).

3.4.3. Multiple independent acquisitions of *pvl* occurred during the evolution of CC30 pandemic clones

A BLAST search to identify the *pvl* locus (lukS-PV and lukF-PV) in the genomes of the CC30 isolates included in the current study revealed it to be present in 19 of 21 isolates belonging to the community-associated phage type 80/81 and SWP clones and absent from the EMRSA-16 and other epidemic CC30 clones (Figure 3.2).

The *pvl* loci are encoded by temperate phage that can be differentiated into distinct morphological groups on the basis of their elongated- or icosahedral-head types (Ma et al., 2008). Sequence analysis of the small and large terminase subunits of the PVL phage revealed that 3/4 SWP clone isolates contained the elongated-head phage type, whereas the remaining SWP isolate and the phage type 80/81 isolates had the icosahedral-head phage type. Phylogenetic analysis of the small and large terminase subunit gene products indicates the close relatedness of the icosahedral phage heads of the SWP and phage type 80/81 clones (Figure 3.3), suggesting that a progenitor of the SWP and phage type 80/81 clones, which existed over 100 years ago, had previously acquired an icosahedral-head phage type encoding PVL (although independent horizontal acquisition cannot be ruled out). Subsequently, the phage was replaced by the elongated-head phage type in some isolates of the SWP clone (Figure 3.2).

3.4.4. Identification of polymorphisms that correlate with the nosocomial habitat of EMRSA-16

Genome-wide analysis of the CC30 sequences indicates that 58 of 60 EMRSA-16 isolates contained the type II staphylococcal cassette chromosome mec element (SCC*mec*II), and that this cassette is restricted to the EMRSA-16 lineage (Figure 3.2).

In addition, non-synonymous mutations were identified in loci implicated in resistance, including genes encoding penicillin binding proteins 2 and 4, the *vraD* component of the bacitracin resistance pathway, and the gene encoding the multidrug efflux transporter NorA in 50 of 60 EMRSA-16 isolates. Specifically, a S84L

replacement in DNA gyrase subunit A, and an S80F replacement in DNA topoisomerase IV subunit A have been demonstrated to confer resistance to fluoroquinolones (Banerjee et al., 2010; Couto et al., 2008; Hiron et al., 2011; Łeski et al., 2005). Of note, a selection of 16 EMRSA-16 isolates tested on the Vitek 2 system (bioMérieux, Basingstoke, UK) were all fluoroquinolone resistant (data not shown). Finally, plasmids encoding resistance to quaternary ammonium compounds were identified in 7 of 60 EMRSA-16 isolates.

In addition to polymorphisms impacting on antibiotic sensitivity, mutations were observed in the virulence determinants *agr* and *hla* previously identified in an ancestor of the EMRSA-16 clone by DeLeo and colleagues (DeLeo et al., 2011). The *agr* and *hla* polymorphisms were demonstrated to result in reduced virulence phenotypes in murine models of infection (DeLeo et al., 2011). In the current study these polymorphisms were identified in the hospital restricted EMRSA-16 clade and in strains belonging to the sister clade, annotated as other epidemic CC30. Furthermore, a nonsense mutation occurred in the squalene desaturase gene (*crtM*) leading to pseudogene formation and disruption of the terminal portion of the mevalonate pathway, which leads to staphyloxanthin carotenoid biosynthesis (Liu et al., 2005). This is consistent with the non-pigmented colony phenotype of strains described in the earliest reports of the EMRSA-16 clone (CDRW, 1997).
In addition, non-synonymous mutations were identified in genes encoding proteins involved in virulence including the virulence gene regulator CcpA and, for a proportion of EMRSA-16 strains, cell wall-associated proteins Fib, SasH, IsdB, and Ebh.

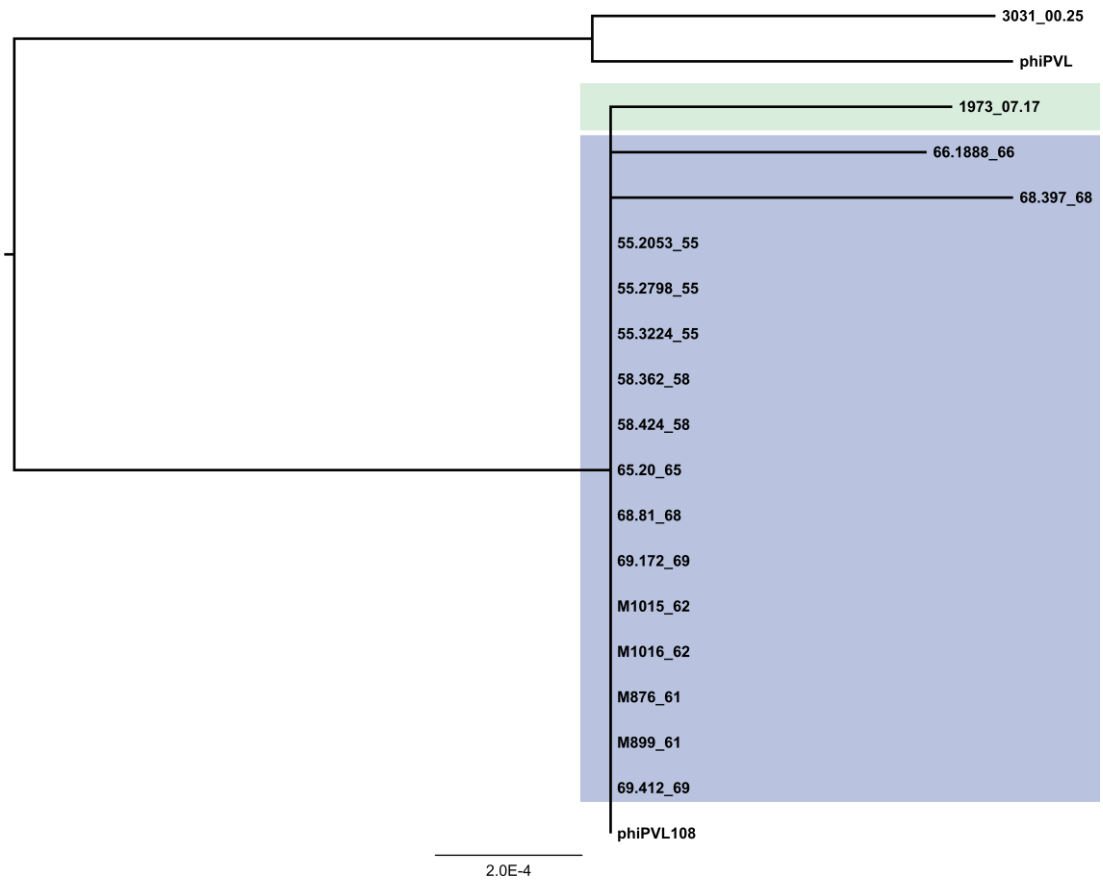


Figure 3.3 Phage type 80/81 and southwest Pacific strains harbour closely related PVL phage with icosahedral head morphology. Neighbour-joining tree of coding sequence for terminase large and small subunits for CC30 isolates containing the icosahedral-head phage type. Blue and green shading denotes isolates belonging to 80/81 and southwest Pacific clone respectively. Unshaded branches represent other PVL phage with icosahedral head type. Scale bar represents substitutions per site.

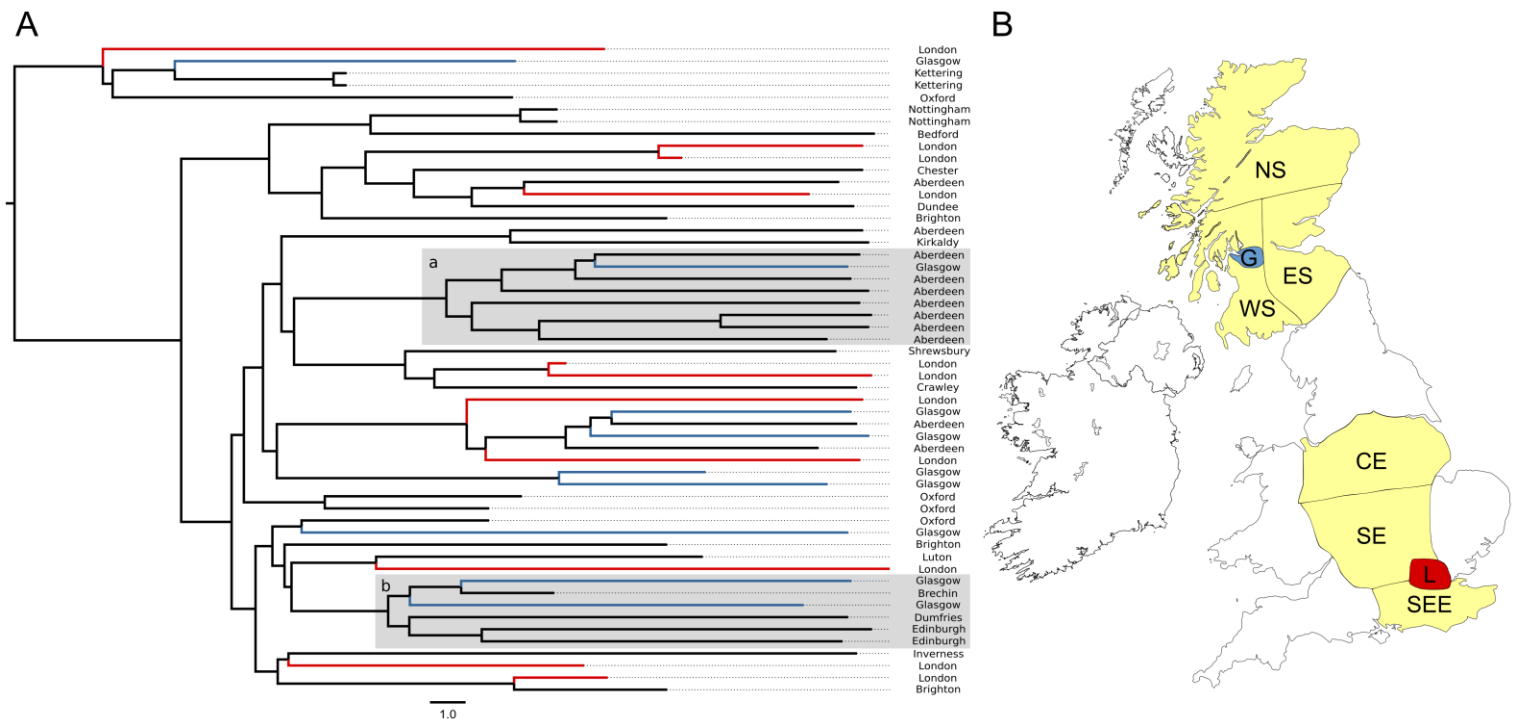


Figure 3.4 Large population centres are hubs for transmission of EMRSA-16. (A) Bayesian phylogenetic reconstruction of UK EMRSA-16 isolates. Terminal branches representing London and Glasgow isolates are coloured red and blue respectively. Black branches represent isolates from other UK locations. Branches are scaled with time, and the scale bar represents 1 year. Grey shading indicates examples of geographically restricted subclades a and b. (B) Map of he UK indicating regions represented in the phylogeographic analysis (Coloured yellow, red, and blue). Isolates were grouped by geographic region; L, London; SEE, southeast England; SE, south England; CE, central England; NS, north Scotland; ES, east Scotland; WS, west Scotland). Blue and red shaded regions represent Glasgow (G) and London (L) respectively.

3.4.5. Phylogeographic analysis of UK EMRSA-16 sequences suggests that large populations centres are hubs for transmission

Bayesian phylogenetic analysis of a subset of the genome sequence data, which included the UK EMRSA-16 isolates only (54 of 60 EMRSA-16 isolates), resolved several subclades, consistent with the existence of EMRSA-16 strains that are endemic to particular hospitals or regions (Figure 3.4a). In particular, EMRSA-16 strains isolated in Aberdeen Royal Infirmary between 2006 and 2007 are more closely related to each other (subclade a) than to other EMRSA-16 isolates (Figure 3.4a). In addition, subclade b consists largely of isolates from central Scotland, implying the existence of an EMRSA-16 subtype, which is endemic to this region (Figure 3.4a). However, isolates from London and Glasgow are widely distributed among clusters of closely related isolates from regional hospitals, consistent with hospitals in major population centres acting as reservoirs for EMRSA-16 UK transmissions. To examine this observation further, the discrete phylogeographic diffusion model implemented in BEAST was used, with isolates grouped by geographic region (London, southeast, south, and central England and north, east, and west Scotland (Figure 3.4b)). Using the regional groupings, statistical support was identified for transmission events of EMRSA-16 between London and the south and southeast regions of England (Bayes factors 3.00 and 4.36, respectively). In addition, Glasgow (west Scotland) was identified as a reservoir for transmission of EMRSA-16 to surrounding population centres in the north and east of Scotland (Bayes factors 4.04 and 7.16, respectively).

3.4.6. A single acquisition of the *tst* gene led to CC30 strains responsible for the toxic shock syndrome epidemic

Previous studies have demonstrated that the majority of cases of menstrual TSS are caused by a single clone that corresponds to CC30 (Fitzgerald et al., 2001; Peacock et al., 2002). However, the distribution of the *tst* gene encoding the toxic shock syndrome toxin 1 (TSST-1) among CC30 subclades has not been previously examined. As the *tst* gene is not present in the MRSA252 reference strain, contigs assembled *de novo* were interrogated using BLAST for presence of the *tst* locus. Sequence analysis determined that the *tst* gene in these strains is harboured by the staphylococcal pathogenicity island 2 (SaPI2), and is present in 55 of 66 (83%) of the isolates in the sister clades represented by EMRSA-16 and other CC30 epidemic isolates (Figure 3.2), but is not present in any of the 21 (0%) isolates from the clades represented by the SWP and 80/81 clones. These data imply the restriction of the *tst* gene to specific contemporary CC30 clones and its absence from the phage 80/81 and SWP clones.

The high level of sequence identity of SaPI2 among the 55 *tst* positive isolates (5 or fewer polymorphic sites in 14.7-kb SaPI2) strongly suggests that a single acquisition event occurred in an ancestor that existed 10–140 years before the TSS epidemic of the 1970/80s (Figure 3.2).

3.4.7. Extended population analysis of the origin of the TSS epidemic

In order to further investigate the lineage restriction of TSST-1, an additional 67 CC30 sequences were selected for whole genome sequencing. In particular the

diversity and evolutionary history of *S. aureus* strains associated with TSS was examined. Strains were selected based on the results of MLST, PFGE and PCR screening for the presence of *tst*. Additionally, isolates from known cases of TSS (Fitzgerald et al., 2001), or TSST-1 related food poisoning were included. In total 117/154 isolates positive for *tst* were included, with isolation dates ranging from 1964 to 2012 (Appendix 1).

A maximum-likelihood phylogenetic analysis built using the core genome sequence resulted in a well-supported tree, reconstructed using a core genome alignment of 1,803,190 bp with 9,510 variable sites (Figure 3.5). The majority of the newly sequenced isolates clustered with the strains defined as other epidemic CC30 (42/67), 10/67 belonged to the SWP clone, and 2/67 were positioned basally within the EMRSA-16 lineage. 5/67 were identified as ST39, a double locus variant of both ST30 and ST36, and formed a distinct clade within the CC30 phylogeny. The remaining 8 sequenced isolates were positioned on long branches with ancestral nodes basal to the ST39 and other epidemic clades.

Of the 67 additional isolates, 46 were positive for the *tst* gene. Sequence analysis of the integrase gene of the associated SaPI suggests 4 independent acquisitions of TSST-1 by CC30 strains (Figure 3.6), but a single diverse clade of community-associated CC30 strains was responsible for the majority of the TSS epidemic of the late 1970s, and for subsequent cases of TSS in the UK and USA (Figure 3.5). There was an additional acquisition of a TSST-1 carrying SaPI on the branch leading to the ST39 clade (yellow), and 2 separate acquisitions by single locus variants of ST30.

Of note, within the previously defined EMRSA-16 clade, 2 isolates from known TSS infections were situated basally (Figure 3.5).

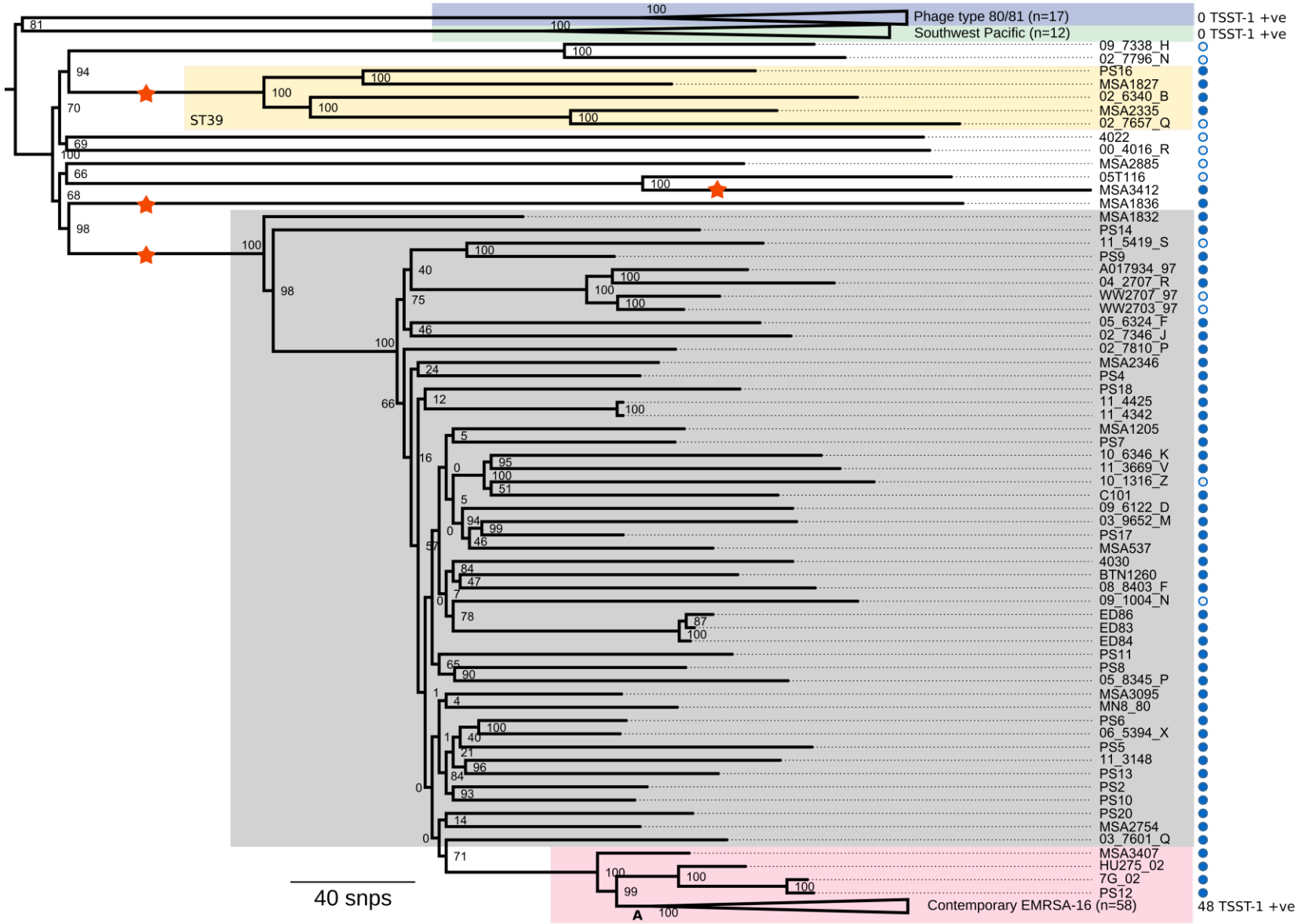

Phage type 80/81 (n=17)
Southwest Pacific (n=12)
ST39
Contemporary EMRSA-16 (n=58)
A
40 snps
0 TSST-1 +ve
0 TSST-1 +ve
48 TSST-1 +ve
09_7338_H
02_7796_N
PS16
MSA1827
02_6340_B
MSA2335
02_7657_Q
4022
00_4016_R
MSA2885
05T116
MSA3412
MSA1836
MSA1832
PS14
11_5419_S
PS9
A017934_97
04_2707_R
WW2707_97
WW2703_97
05_6324_F
02_7346_J
02_7810_P
MSA2346
PS4
PS18
11_4425
11_4342
MSA1205
PS7
10_6346_K
11_3669_V
10_1316_Z
C101
09_6122_D
03_9652_M
PS17
MSA537
4030
BTN1260
08_8403_F
09_1004_N
ED86
ED83
ED84
PS11
PS8
05_8345_P
MSA3095
MN8_80
PS6
06_5394_X
PS5
11_3148
PS13
PS2
PS10
PS20
MSA2754
03_7601_Q
MSA3407
HU275_02
7G_02
PS12

Figure 3.5 Maximum-likelihood phylogenetic reconstruction of the CC30 lineage using all sites in core genome. Blue, green, red, grey, and yellow shading correspond to the 80/81, southwest Pacific, EMRSA-16, toxic shock epidemic, and ST39 CC30 clones respectively. Presence of the *tst* locus is indicated by shaded blue circles, and acquisitions of SaPIs carrying the *tst* locus are indicated by stars. 80/81, southwest Pacific, and contemporary EMRSA-16 clades are collapsed at the most recent common ancestor of each clade. Tips are labelled with isolate names, and bootstrap support is shown on nodes.

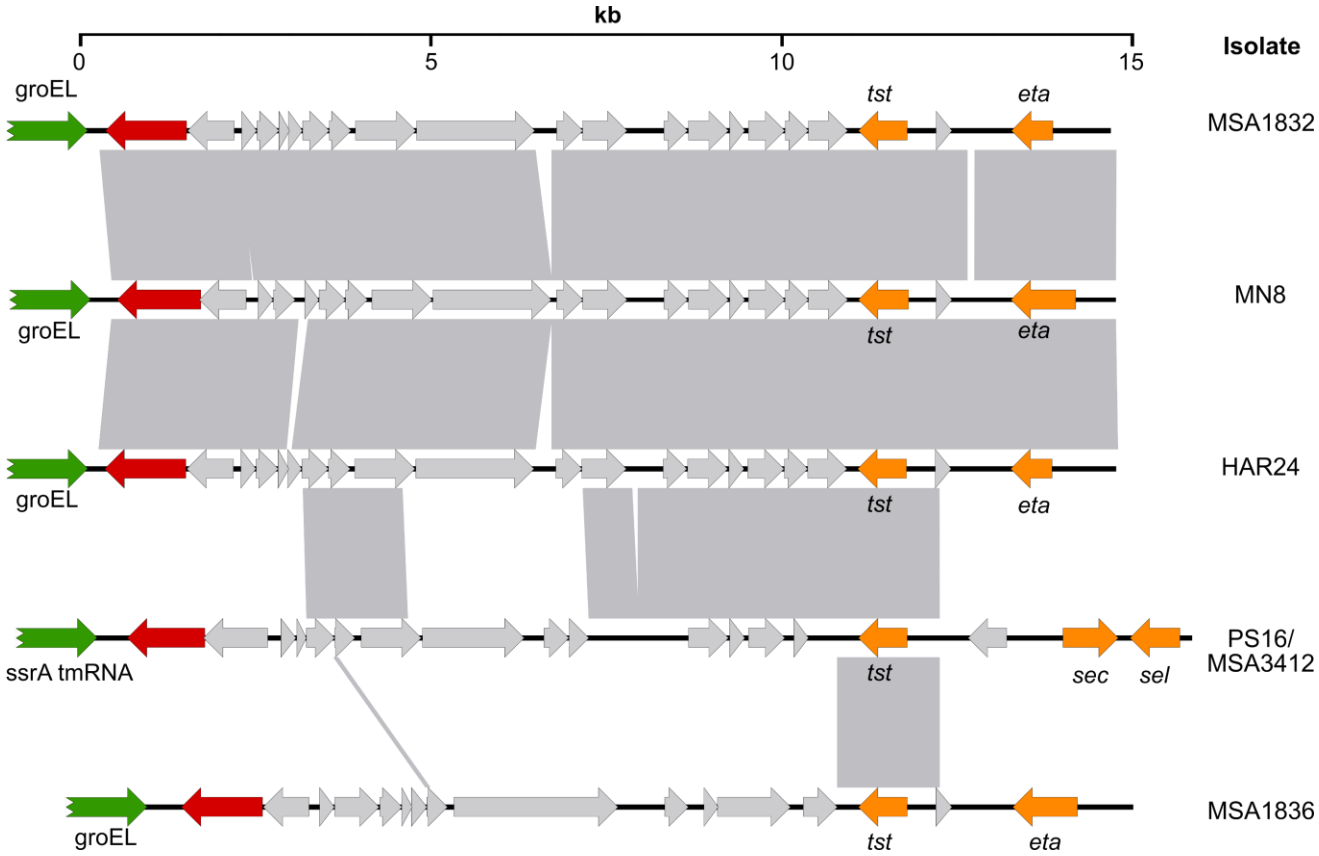


Figure 3.6 *Staphylococcus aureus* CC30 isolates contain 3 distinct pathogenicity islands (SaPIs) encoding the *tst* locus. Predicted coding sequences are indicated with shaded arrows; integrase, toxin, and chromosomal genes are shaded red, orange, and green respectively. Regions of nucleotide identity (>85%) between SaPIs are highlighted with grey shading. Isolates MSA 1832 and MN8 are from known cases of TSS, HAR24, PS16, and MSA 1836, and MSA 3412 are representative of ST36, ST39, ST442, and ST37 respectively.

3.4.8. Expression levels of TSST-1 and resistance to oxacillin are variable among *tst*+ CC30 isolates

Differences in expression levels of TSST-1 by isolates representative of different clades within the CC30 phylogeny were assessed by Western immunoblotting (work performed by Robyn A. Cartwright, Figure 3.7). Isolates from known cases of TSS displayed high levels of expression, with the exception of a single strain (PS12) that is of ST500 and is positioned basal to the EMRSA-16 clade in the CC30 phylogenetic tree (Figure 3.5). Isolates of the ST36 displayed low levels of TSST-1 expression, with the exception of the most basal strain (MSA 3407), which displayed a high level of expression.

Differences in the minimum inhibitory concentration (MIC) of oxacillin for a representative sample of EMRSA-16 isolates was assessed using the Etest method (bioMérieux). The EMRSA-16 isolates displayed a high level of resistance to oxacillin (MIC > 256 μg/ml), with the exception of the most basal isolate, MSA 3407 (Figure 3.5), which had a 4-fold lower MIC of 64 μg/ml (Figure 3.7) (work performed by Robyn A. Cartwright).

Taken together, these observations suggest that progenitors of the EMRSA-16 lineage were capable of causing TSS, and that subsequent reduction of TSST-1 expression, together with an increased level of resistance to oxacillin may account for the adaptation of the clone to the hospital environment.

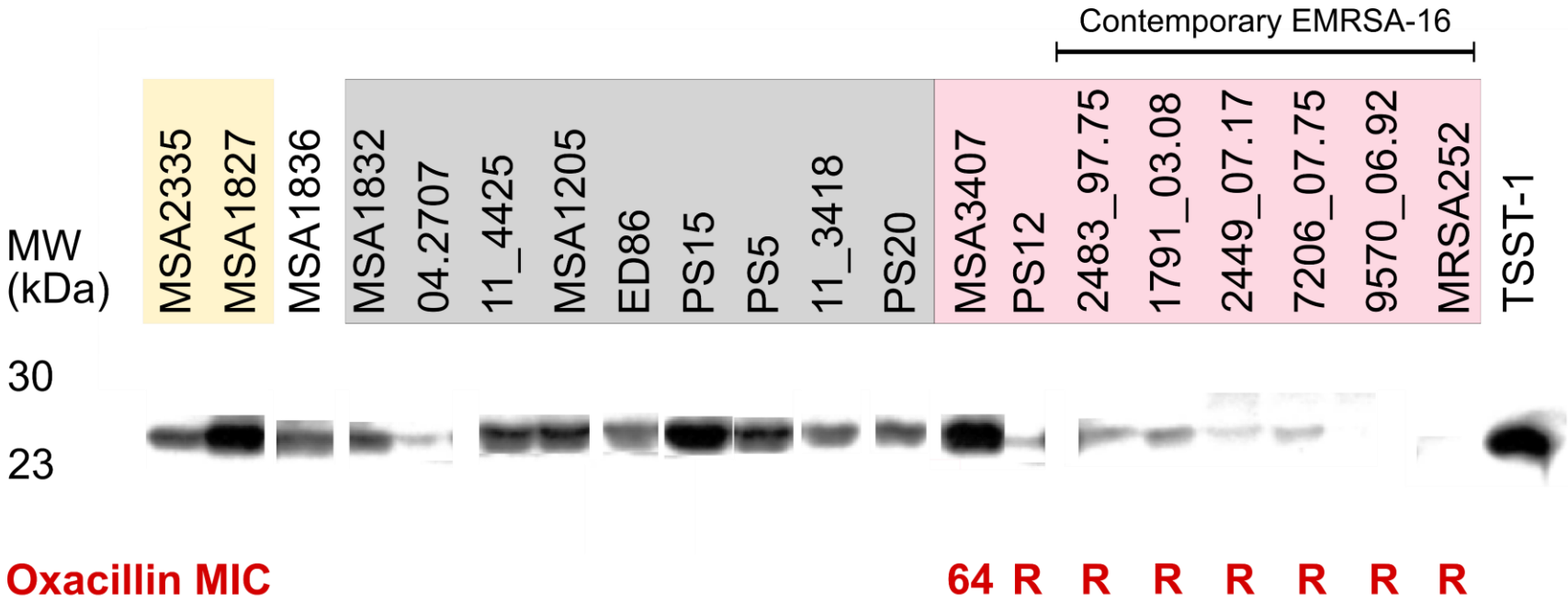


Figure 3.7 Expression levels of TSST-1 are variable across CC30 isolates that carry the *tst* locus. Concentrated supernatants expressing TSST-1, resolved on SDS-PAGE gel. Yellow, grey, and red shaded isolates are representative of ST39, TSS-epidemic, and EMRSA-16 isolates respectively, non-shaded isolates do not belong to an identifiable clade. MRSA252 is included as a negative control, purified TSST-1 is included as a positive control. The oxacillin MIC is shown for EMRSA-16 isolates, R indicates an MIC > 256 µg/ml.

3.4.9. Identification of polymorphisms coinciding with reduced capability to cause TSS

In order to examine the potential molecular correlates of TSS infection, basal TSS associated isolates within the EMRSA-16 lineage were compared with hospital-associated isolates. A total of 26 core-genome polymorphisms were identified on branch A (Figure 3.5). Of note, 4 non-synonymous substitutions were observed within the extracellular L2 loop of the penicillin binding protein encoded by *blaR1*. The L2 loop forms an intra-molecular interaction with the C-terminal β-lactam receptor, which when disrupted results in transcription of *blaZ* and *mecA*, and increased levels of resistance to β-lactam antibiotics (Figure 3.8) (Hackbarth et al., 1993; Hanique et al., 2004; Thumanu, 2006). The polymorphisms within *blaR1* could have a dual effect on the antibiotic resistance phenotype and virulence potential of hospital-associated clones. Expression of the *mecA* encoded PBP2a represses *agr* activity, leading to reduced expression of virulence factors, including TSST-1 (Figure 3.8) (Pozzi et al., 2012; Recsei et al., 1986; Rudkin et al., 2012).

In addition, a non-synonymous polymorphism was identified in the virulence gene regulator *ccpA*. Isolates from TSS cases, and the most basal EMRSA-16 isolate, MSA 3407 (isolates with high levels of TSST-1 expression), had the same sequence at the *ccpA* locus, while the remainder of the EMRSA-16 isolates shared an allele with the amino acid altering polymorphism. Like *blaR1*, *ccpA* is a regulator of TSST-1 expression (Figure 3.8) (Seidl et al., 2008), and *ccpA* mutants have been shown to have reduced levels of resistance to oxacillin (Seidl et al., 2006).

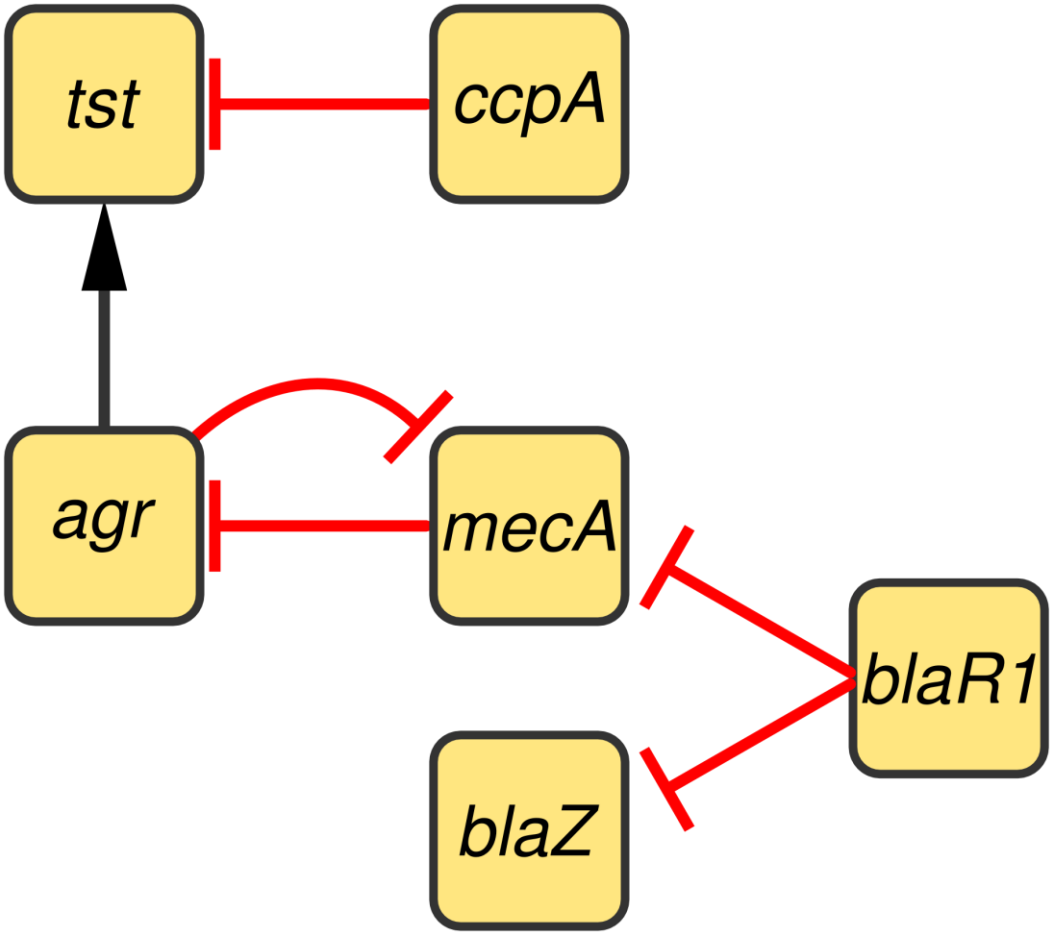


Figure 3.8 Regulation of expression of virulence factor and antibiotic resistance loci by *ccpA* and *blaR1.* Transcription of the *tst* locus is directly regulated by *ccpA*, and indirectly through *blaR1*. Resistance to β-lactam antibiotics is mediated through *blaR1* regulation of *mecA* and *blaZ.* Black arrows represent upregulation, and red bars represent downregulation.

3.5. Discussion

S. aureus CC30 is a globally disseminated lineage that has had an important impact on human health over a number of decades (Altemeier et al., 1982; Nahmias et al., 1961; Robinson et al., 2005a). The 80/81 and SWP clones have been responsible for severe community associated infections, while the EMRSA-16 clone is endemic in UK hospitals causing significant morbidity and mortality. Previously, Robinson and colleagues used a combination of MLST and PCR genotyping to examine the evolution of CC30 pandemic clones and inferred that the SWP clone originated from the historic phage type 80/81 clone (Robinson et al., 2005a). However, an important recent study by DeLeo et al. based on a comparative genome sequence analysis of 9 CC30 isolates demonstrated that the SWP clone is not a direct descendent of the phage 80/81 clone, but that each clone has evolved from a single ancestral strain, which gave rise to all 3 CC30 pandemics (DeLeo et al., 2011). Consistent with these findings, the phylogenetic analysis of the current dataset clearly demonstrates the independent origins of each of the three major CC30 pandemic clones.

Furthermore, the number of genome sequences included in this study facilitated the use of a Bayesian coalescent method to estimate the rate of molecular evolution within CC30, the date of emergence of each of the major pandemic clones, and the timing of other important CC30 evolutionary events. The estimate of the evolutionary rate reported here is consistent with those reported for other *S. aureus* clonal complexes (Harris et al., 2010; Holden et al., 2013; Nübel et al., 2010). Calibration of the molecular clock allows the date of emergence of each of the major pandemic clones, and the timing of other important CC30 evolutionary events to be estimated. Temporal analysis of bacterial epidemics may implicate

contemporaneous environmental factors or human practices that promoted their emergence and expansion. For example, the phage type 80/81 clone is predicted to have emerged shortly before a time of intensive penicillin use, which may have provided a selection for clonal expansion after acquisition of a β-lactamase plasmid (Gould, 1958; Robinson et al., 2005a; Rountree et al., 1961). For each of the CC30 pandemic clones, the predicted date of the MRCA is several years earlier than the time that these clones were first reported as clinical isolates in the literature, consistent with observations for the pandemic ST239 clone of *S. aureus* (Gray et al., 2011; Harris et al., 2010; Smyth et al., 2010). Early strains of the emergent clones may have existed for some time before acquiring mobile genetic elements, or before encountering selective pressures, which promoted rapid clonal expansion. Furthermore, it is feasible that EMRSA-16 strains were causing low-level clinical infections for several years before they were first described in the literature as a significant clinical problem.

Comparative genomic analysis of closely related isolates from epidemics occupying different niches or associated with different disease manifestations is a powerful means for identifying genetic events that may have contributed to clone emergence and pathogenesis (Croucher et al., 2011; Harris et al., 2010; Hendriksen et al., 2011; Mutreja et al., 2011; Price et al., 2012). PVL has been a marker for community-associated clones of *S. aureus* associated with skin and soft tissue infections or severe necrotizing pneumonia (Diep et al., 2010), although increased identification of PVL positive strains associated with nosocomial infections weakens this correlation (Popovich et al., 2008; Rossney et al., 2007). The finding that the PVL

phage harboured by the phage type 80/81 clone shares a head morphology with an isolate from the SWP clone suggests that PVL has a long residency with some community-associated *S. aureus* clones, having been maintained in CC30 clades since a likely acquisition event that occurred about 137 y ago (95% HPDs, 86–197 y ago). Consistent with these findings, others have described SWP and phage type 80/81 strains harbouring similar PVL phage types (Chen et al., 2013). These data are consistent with a central role for PVL in the success of some community-associated *S. aureus* clones (Diep et al., 2004).

Since the major TSS epidemic among menstruating women began in 1978, the incidence has fallen from 13.7/100,000 population to 0.52/100,000 population in the years 2000-2003, but TSS remains a important clinical problem (DeVries et al., 2011). The current study demonstrates that carriage of the *tst* gene is restricted to defined sublineages within the CC30 population, representative of both community-associated and hospital-associated strains. Of note, a large proportion of strains belonging to the hospital-associated EMRSA-16 clone encode TSST-1, but these strains are not associated with TSS infections. The low level of sequence diversity within SaPI2, the pathogenicity island that harbours TSST-1, is suggestive of a single acquisition of SaPI2 prior to clonal expansion of the strains responsible for the TSS epidemic, and EMRSA-16. This interpretation supports the earlier finding based on electrophoretic typing that the TSS epidemic was caused by already widely disseminated TSST-1 positive strains rather than rapid clonal expansion of a single strain that had acquired a fitness advantage (e.g., SaPI2 acquisition) (Fitzgerald et al., 2001). The restricted distribution of TSST-1 across the CC30 lineage, and the

finding that there are limited introductions prior to clonal expansion, further supports the pre-adapted clone hypothesis.

It remains to be determined why this clone continues to harbour the *tst* locus, but the presence of TSST-1 in a large proportion of contemporary isolates from the EMRSA-16 clone may suggest another function in addition to the role of TSST-1 as a superantigen. In addition, SaPI2 may harbour a yet to be determined locus that confers an advantage to EMRSA-16 strains in the hospital setting. This study shows that CC30 has acquired TSST-1 on at least 4 occasions, but a single diverse clade of community-associated CC30 strains was responsible for the TSS epidemic of the late 1970s, and for subsequent cases of TSS in the UK and North America. Phenotypic analysis of TSST-1 identified considerable variation in expression levels among CC30 clades, which may influence the capacity to cause TSS in the hospital setting with the majority of EMRSA-16 isolates tested displaying low levels of TSST-1 expression. High-resolution genomic analysis revealed genetic determinants that may underlie this variation, specifically a non-synonymous polymorphism in ccpA, together with 4 non-synonymous mutations present in the extracellular L2 loop of *blaR1* of contemporary EMRSA-16 strains and their immediate ancestral strains. These polymorphisms have the potential to alter the expression of PBP2a through regulation of *mecA* transcription (Figure 3.8) (Hanique et al., 2004; Thumanu, 2006), which in turn could lead to a reduction in TSST-1 expression through repression of the *agr* system (Recsei et al., 1986; Rudkin et al., 2012), although this hypothesis warrants further investigation. Overall, these data resolve the evolutionary history of

TSS caused by *S. aureus* and reveal novel genetic determinants that may profoundly influence pathogenic potential.

The hospital environment provides a unique array of insults that bacteria must overcome, including numerous classes of antibiotics used in treatment and routine use of disinfectants and antiseptics. However, in contrast to the community setting, the hospital provides a continuous supply of immune-compromised hosts, who offer plentiful opportunities for infection and transmission. Given the strong selective pressures encountered in the nosocomial environment, it is unsurprising that the EMRSA-16 clone carries numerous antibiotic resistance determinants. Acquiring resistance to fluoroquinolones has been shown to have been an important event in the evolution of the endemic UK hospital-associated EMRSA-15 clone (Holden et al., 2013), and likely contributes to the ability of EMRSA-16 to survive in the nosocomial setting. Further differences in antibiotic sensitivity between countries suggests that differing antimicrobial prescribing practices select for the emergence of resistance at a local level (Holden et al., 2013).

Acquisition of SCC*mec*II was a critical genetic event in the evolution of the EMRSA-16 clone as a hospital-associated antibiotic-resistant clone refractory to treatment with β-lactam antibiotics. Of note, the type II SCC*mec* element has been demonstrated to reduce the toxicity of MRSA CC30 strains in comparison with methicillin sensitive CC30 strains (Collins et al., 2010), and prevents normal stationary phase induction of the *agr* system, leading to decreased expression of cytolytic toxins (Rudkin et al., 2012). It is speculated that this reduction in energy

requirement could compensate for the metabolically costly maintenance of a large SCC*mec* element and its associated methicillin resistance, but which would likely lead to reduced fitness outside of the hospital setting. Recently, the carriage of the *psm-mec* locus on type II and III SCC*mec*, has been demonstrated to reduce the expression of the PSMα toxin through repression of the *agr* system (Kaito et al., 2013), although this effect was not replicated in an EMRSA-16 strain, suggesting that the reduced virulence effect of *psm-mec* carriage is dependent on the genetic background of a strain (Kaito et al., 2013; Queck et al., 2009).

In contrast, the EMRSA-15 clone emerged following the single acquisition of SCC*mec*IV, a smaller cassette that encodes fewer resistance determinants, and is capable of causing infection in the community (Espadinha et al., 2013; Holden et al., 2013). The level of cytolytic toxin expression of strains carrying a type IV SCC*mec* element is comparable to isogenic strains with the SCC*mec* element deleted (Collins et al., 2010), which may help to explain the success of the clone in the community. As proposed by the DeLeo et al., mutations in *agr* and *hla* may have contributed to the hospital association of contemporary CC30 isolates (DeLeo et al., 2011). However, isolates that belong to the sister clade to EMRSA-16, annotated as other epidemic CC30 contain the same *agr* and *hla* mutations but have been isolated from episodes of severe community-associated infections of healthy humans such as TSS (Figure 3.2) (Subedi et al., 2007). Therefore, while *agr* and *hla* mutations are indeed likely to influence the capacity of CC30 isolates to cause certain types of infection, they are probably not sufficient to explain the hospital restriction of the EMRSA-16 clone since its differentiation from the other epidemic CC30 clade.

Additional polymorphisms are predicted to have an impact on the virulence potential of EMRSA-16. For example, the formation of a *crtM* pseudogene, leads to disruption of the staphyloxanthin biosynthesis pathway. Bacterial carotenoids confer resistance to oxidative killing during phagocytosis, and staphyloxanthin-deficient *S. aureus* mutants have diminished virulence in animal models of infection (Liu et al., 2005). The loss of metabolically expensive pigment production may in part compensate for the energy cost of maintaining the large SCC*mec*II element mediating antibiotic resistance. In turn, and in combination with the previously described *hla* and *agr* mutations, lack of pigment may affect its capacity to cause disease of healthy humans outside of the hospital setting. There are a number of other polymorphisms in loci that could be implicated in affecting virulence potential, such as in the virulence gene regulator CcpA in, the cell wall-associated proteins Fib, SasH, IsdB, and Ebh. Similar polymorphisms are observed in cell wall-associated proteins in the EMRSA-15 lineage, with a nonsense mutation causing the formation of a Ebh pseudogene, and a fusion of the FnbPA and FnbPB CDSs (Holden et al., 2013). It should be pointed out that some of the mutations are likely to be the result of fixation due to genetic drift before clonal expansion rather than the result of selective pressures encountered in hospitals.

Whereas the examples discussed are selected on the basis of known or implicated roles in virulence, the possibility that some of the other non-synonymous mutations identified among EMRSA-16 strains may also have had a role in shaping its hospital-specialist lifestyle cannot be dismissed. In contrast to DeLeo et al. who neatly demonstrated the virulence-attenuating effects of mutations in *agr* and *hla* (DeLeo et

al., 2011), the functional consequences of the polymorphisms described here have not yet been elucidated. However, the identification of EMRSA-16 clade-specific mutations provides important avenues for future studies examining the hospital specialization of this important clone.

The geographic spread of hospital-associated bacterial clones is not well understood. Whole genome sequencing of large numbers of nosocomial isolates allows the high-resolution tracking of the transmission of strains through space and time (Beres et al., 2010; Mutreja et al., 2011; Okoro et al., 2012). Phylogeographic analysis of the EMRSA-15 clone revealed a likely origin in central England, with restriction to this region until the late 1980s, followed by repeated transmission events via multiple routes throughout the UK and Europe (Holden et al., 2013). The predicted timings of emergence for a given region were concordant with the first clinical reports of EMRSA-15 (Holden et al., 2013). The phylogeographic analysis of the UK EMRSA-16 strains provide evidence for transmission routes from hospitals in major cities to UK regions, leading to endemic strains circulating in local hospitals. These findings are consistent with a recent US study that estimated high transmission rates between large hospitals and long-term care facilities using a simulation model (Barnes et al., 2011). It has previously been postulated that an increase in the willingness of patients to travel further for treatment, coupled with the centralization of specialist treatment centres have been contributing factors to the spread of MRSA throughout the United Kingdom (Murchan et al., 2004). These data could inform the design of infection control protocols, such as decolonization of patients before transfer from large hospitals, to limit inter-hospital transmission as a major driving

force for epidemics. The number of isolates included, and the relatively small number of hospitals sampled limits the analysis, as large regions of the United Kingdom are not represented in the dataset. For example, it would be interesting to examine the transmission dynamics between hospitals within a single healthcare trust, or to integrate epidemiological data such patient transfer and bed occupancy rates into the phylogeographic analysis.

Recently, the prevalence of EMRSA-16 has declined within UK hospitals and this has variously been attributed to improved infection control programs, phage epidemics, and changes in prescribing habits (Ellington et al., 2010; Knight et al., 2012; Wyllie et al., 2011). Integrating genomic data with these observations could offer an insight into the molecular correlates underpinning the decline, and lead to improved infection control programs in future.

Increasingly, the boundaries between the community and hospital are becoming blurred, and lineages are emerging capable of causing pathology in both settings (Espadinha et al., 2013; Popovich et al., 2008; Rossney et al., 2007). Identification of transmission pathways will be key to controlling the spread of these strains, and WGS offers a rapid and proven way of determining pathogen spread (Harris et al., 2013; Lieberman et al., 2011).

3.6. Conclusions

The current study builds on recent findings that revealed the independent origin of phage 80/81 and SWP pandemics and identified mutations leading to reduced

virulence among contemporary CC30 strains (DeLeo et al., 2011). The capacity to rapidly sequence the genome of large numbers of bacterial isolates can result in enhanced insights into bacterial epidemics. By using a high-resolution phylogenomic approach, a timeframe has been determined for the emergence and expansion of CC30 pandemics, and identified pathways of transmission of a major hospital-associated MRSA clone, which may be used to inform control methods. In addition, unique genetic events that correlate with its adaptation to the hospital niche were identified, some of which may help to explain its lack of success in the community setting. The use of a similar approach for other hospital-associated bacteria could lead to the identification of risk factors that promote the emergence and expansion of epidemics, and thereby inform the rational design of methods for controlling their inter- and intra-hospital spread.

4. Molecular epidemiology of an outbreak of Legionnaires' disease in Edinburgh, Scotland

4.1. Introduction

Legionella pneumophila is ubiquitous in both natural and manmade freshwater reservoirs (Brown et al., 1999; Fliermans et al., 1979; Morris et al., 1979; Sanchez et al., 2008; Thomas et al., 1993), and at low doses does not pose a risk to human health. Aerosolisation of *L. pneumophila* due to increased temperatures, vapour pressure, and bacterial density in a reservoir result in increased exposure of the human population to the pathogen, leading to an elevated risk of developing legionellosis (Brown et al., 1999; Conza et al., 2013). Male gender, older age, smoking and underlying respiratory pathology are associated with increased risk of developing the severe form of legionellosis, Legionnaires' disease (LD) (Marston et al., 1994).

In Scotland, the incidence of disease associated with *L. pneumophila* ranges between 15 and 40 cases per year, with approximately two-thirds of cases attributed to infection during travel (Potts et al., 2011). Between 31st May and 17th July 2012, a total of 56 confirmed and 35 suspected cases of LD were reported with an epidemiological link to the south-west region of Edinburgh, Scotland. Analysis of meteorological conditions preceding and during the outbreak suggested a cluster of cooling towers in the north-east of the affected area as the likely source of the outbreak (McCormick et al., 2012). Patients were linked to the outbreak based on the results of traditional typing methods for *L. pneumophila* of sequence based typing (SBT), serotyping, and monoclonal antibody (mAb) subgrouping (McCormick et al., 2012). However, traditional typing methods have limited potential for resolving the source of an outbreak in situations where a genotype is

present in multiple potential sources. Previously, an outbreak of *L. pneumophila* in Hampshire resulted in a total of 26 cases of LD over a 3 week time period in 2003. Traditional typing methods implicated 2 cooling towers on a single site as the likely source of the outbreak (Reuter et al., 2013). The retrospective application of whole genome sequencing of 2 outbreak strains and 4 environmental strains isolated at the time of the outbreak offered insights into the relationship between the isolates. A close phylogenetic relationship was identified between isolates from the cooling towers and isolates from patients. However, an isolate obtained from a domestic spa pool also clustered with the outbreak strains, meaning this site cannot be ruled out as a source of the outbreak. An isolate from a cooling tower on a separate site was phylogenetically distant to the outbreak strains and had previously been eliminated as being the likely source due to having a monoclonal antibody subgroup type distinct from the outbreak strains.

In the current study, whole genome sequencing was utilised to explore the diversity of *L. pneumophila* isolates from individuals within the Edinburgh outbreak. Application of phylogenetic and comparative genomic techniques revealed an unexpectedly high level of genomic diversity among the outbreak strains, with important implications for future investigations of infectious disease outbreaks.

4.2. Aims

- To understand the levels of diversity present in isolates related to an outbreak.
- To evaluate the utility of whole genome sequencing over traditional typing methods for investigating *L. pneumophila* outbreaks.

4.3. Methods

4.3.1. Identification of LD

Cases of LD are defined as a patient having a clinically defined pneumonia together with either isolation of Legionella species from respiratory secretions, detection of *L. pneumophila* antigen in urine, or a positive *L. pneumophila* Sg 1 antibody response. Isolates were successfully cultured from 18/91 patients, and epidemiologically-linked to the outbreak on the basis of isolation date, location, and typing results of Sg 1, SBT of ST191, and a mAb subgroup of Knoxville (Table 4.1). No viable cultures were obtained from water samples of the suspected sources of the outbreak. 5 historical isolates reported as ST191 in the database of the Public Health England (PHE), and 3 contemporaneous but not epidemiologically-linked isolates from Scotland were selected for sequencing, in addition to all cultured isolates from the outbreak (Table 4.1).

4.3.2. *L. pneumophila* culture conditions

L. pneumophila was incubated on buffered charcoal yeast extract (BCYE) agar at 37°C for 48 h in an airtight container with 2 ml water.

4.3.3. *L. pneumophila* genomic DNA extraction and sequencing

A sweep of *L. pneumophila* was taken from a BCYE agar plate cultured for 48 h using a 10 µl loop, suspended in 2 ml PBS, and pelleted by centrifugation for 10 min at 5000 x g. Genomic DNA extraction was performed according to the standard isolation of DNA from Gram-negative bacteria protocol using the QIAcube platform

(QIAgen, Netherlands). Genomic DNA libraries were prepared using the Illumina TruSeq kit, and sequenced using either 150 bp paired end runs on an Illumina MiSeq, or 100 bp paired end runs on an Illumina HiSeq 2000.

4.3.4. *De novo* genome assembly

Processed sequence reads for each isolate were assembled *de novo* using the de Bruijn graph based assembler Velvet v1.1 (Zerbino et al., 2008). Optimal kmer length and coverage cut-off parameters were calculated using the VelvetOptimiser v2.2.5 script (http://bioinformatics.net.au/software.velvetoptimiser.shtml).

4.3.5. Whole genome alignments

Assembled contigs, along with the genome sequences of representative clinical and environmental isolates from public databases (Table 4.2), were aligned using progressiveMauve with default parameters (Darling et al., 2010). Locally collinear blocks (LCBs) not common to all sequences or <1000bp in length were removed, resulting in a gap free core genome alignment.

4.3.6. Recombination detection

To assess the level of recombination among the sequences in the alignment, the gap free genome alignments from progressiveMauve were used as input for BratNextGen (Marttinen et al., 2011). 100 iterations of recombination learning were performed, until parameters had converged, and predicted recombinant regions were removed from the alignment.

4.3.7. Phylogenetic reconstruction

Genome alignments were used as input for maximum likelihood phylogenetic reconstruction in RAxML v7.2.6 (Stamatakis, 2006). A GTR model of nucleotide substitution was applied with the gamma model of rate heterogeneity. Support for nodes was assessed using 1000 bootstrap replicates, and species tree was rooted by using the sequence of a *Legionella longbeachae* isolate as an outgroup.

4.3.8. Genome annotation and identification of variable gene content

Variable gene content among the outbreak isolates was examined using a combination of genome annotation and alignment. Predicted protein and RNA coding sequences in the assembled contigs for the sequenced isolates were annotated using the prokka pipeline v1.5.2 (http://www.vicbioinformatics.com/software.prokka.shtml), and a custom BLAST database of *Legionella* sequences. The Gram negative option was specified to predict signal sequences appropriate for *L. pneumophila*. The annotated assembled contigs were aligned using Mugsy v1r.2.2 (Angiuoli et al., 2011b) and gene presence/absence determined with mugsy-annotator v0.5 (Angiuoli et al., 2011a).

Table 4.1 Origin and typing results of *L. pneumophila* isolates sequenced for this study

Patient	Isolate	Date of isolation	Serogroup	mAb	SBT
1	12_3965	31/05/2012	1	Benidorm	42
2	12_4030	31/05/2012	1	Knoxville	191
	12_4054	31/05/2012	1	Knoxville	191
3	12_4042	01/06/2012	1	Knoxville	191
4	12_4117	02/06/2012	10	NA	1418
5	12_4058	03/06/2012	1	Knoxville	191
6	12_4053	04/06/2012	1	Knoxville	191
7	12_4561	06/06/2012	1	Knoxville	191
8	12_4169	06/06/2012	1	Knoxville	191
9	12_4555	06/06/2012	1	Knoxville	191
10	12_4563	07/06/2012	1	Knoxville	191
11	12_4251	07/06/2012	1	Philadelphia	616

Patient	Isolate	Date of isolation	Serogroup	mAb	SBT
12	12_4499	08/06/2012	1	Knoxville	191
	12_4480	Not recorded	1	Knoxville	191
13	12_5064	08/06/2012	1	Knoxville	191
	12_4437	12/06/2012	1	Knoxville	191
14	12_4240	11/06/2012	1	Knoxville	191
15	12_4903	17/06/2012	1	Knoxville	191
16	12_5223	20/06/2012	1	Knoxville	191
17	12_4904	21/06/2012	1	Philadelphia	37
18	12_5251	29/06/2012	1	Knoxville	191
	12_5392	29/06/2012	1	Knoxville	191
	12_5383	29/06/2012	1	Knoxville	191
	12_5414	29/06/2012	1	Knoxville	191
	12_5415	29/06/2012	1	Knoxville	191
NA	H080160261	2008	6	NA	191

Patient	Isolate	Date of isolation	Serogroup	mAb	SBT
NA	H080160262	2008	6	NA	191
NA	H080160263	2008	6	NA	191
NA	H064020049	2006	1	Allentown	591
NA	H092620872	2009	6	NA	191

mAb, monoclonal antibody subtype; SBT, sequence based typing

Table 4.2 Origin and characteristics of previously sequenced *L. pneumophila* genomes included in this study

Isolate	Isolation date	Isolation country	Sg	mAb	SBT	Source	Reference
Alcoy	1999	Spain	1	ND	578	Clinical	(D'Auria et al., 2010)
Corby	NA	UK	1	Knoxville	51	Clinical	(Glöckner et al., 2008)
ATCC43290	1987	USA	12	NA	187	Clinical	(Amaro et al., 2012)
130b	1978	USA	1	Benidorm	42	Clinical	(Schroeder et al., 2010)
Lens	NA	France	1	Benidorm	15	Clinical	(Cazalet et al., 2004)
Paris	NA	France	1	Philadelphia	1	Clinical	(Gomez-Valero et al., 2011)
Philadelphia	1974	USA	1	Philadelphia	36	Clinical	(Chien et al., 2004b)
NC_018139	NA	France	1	ND	47	Clinical	www.genoscope.cns.fr/agc/microscope
NC_018140	NA	France	1	ND	734	Environmental	www.genoscope.cns.fr/agc/microscope

Sg, serogroup; mAb, mAb, monoclonal antibody subtype; SBT, sequence based typing

4.4. Results

4.4.1. Phylogenetic analysis indicates that the Edinburgh outbreak isolates are genetically distinct from contemporaneous Scottish isolates and other ST191 isolates

In order to determine the level of genetic diversity among isolates from a LD outbreak, the genomes of 22 outbreak-associated *L. pneumophila* isolates were sequenced in addition to 8 isolates not linked to the outbreak (Table 4.1). Sequences were assembled resulting in 26 to 103 contigs per genome, with N50 values ranging from 81,559 to 718,197 bp. In addition, 9 genome sequences from previously published studies were included (Table 4.2).

The ST191 isolates linked to the outbreak shared a core genome of 2,814,826 bp. Phylogenetic reconstruction indicated they were more closely related to each other than to non-outbreak isolates, but formed 2 distinct clades within the species phylogeny (Figure 4.1). There were a total of 344 polymorphisms distributed across the genome that distinguished the 2 clades A and B (Figure 4.2), and therefore the most recent common ancestor is unlikely to have existed in a timeframe commensurate with the duration of the outbreak. Three independent methods for detecting recombination events, BratNextGen (Marttinen et al., 2011), single breakpoint analysis (Kosakovsky Pond et al., 2005), and GARD (Kosakovsky Pond et al., 2006), did not find any evidence of recombination within the core genome alignments (analysis performed by Melissa Ward). Therefore the genetic variation between the 2 clades is unlikely to be the result of recombination. Among clade A

isolates 29 polymorphic sites were identified, while only 4 variable sites were identified in clade B.

The isolate 12_4117 was presumed to be a part of the outbreak based on the results of initial typing at the SHLMPRL, which defined the isolate as Sg1, SBT191 and mAb Knoxville. However, isolate 12_4117 was genetically distinct from all Scottish isolates, and was most related to the *L. pneumophila* strains Alcoy and Corby (Figure 4.1).

The 3 Scottish isolates (12_3965, 12_4251, and 12_4904) not linked to the outbreak were distributed across the tree, and the 4 environmental ST191 isolates were most closely related to the Scottish ST191 outbreak isolates but formed a distinct clade (Figure 4.1).

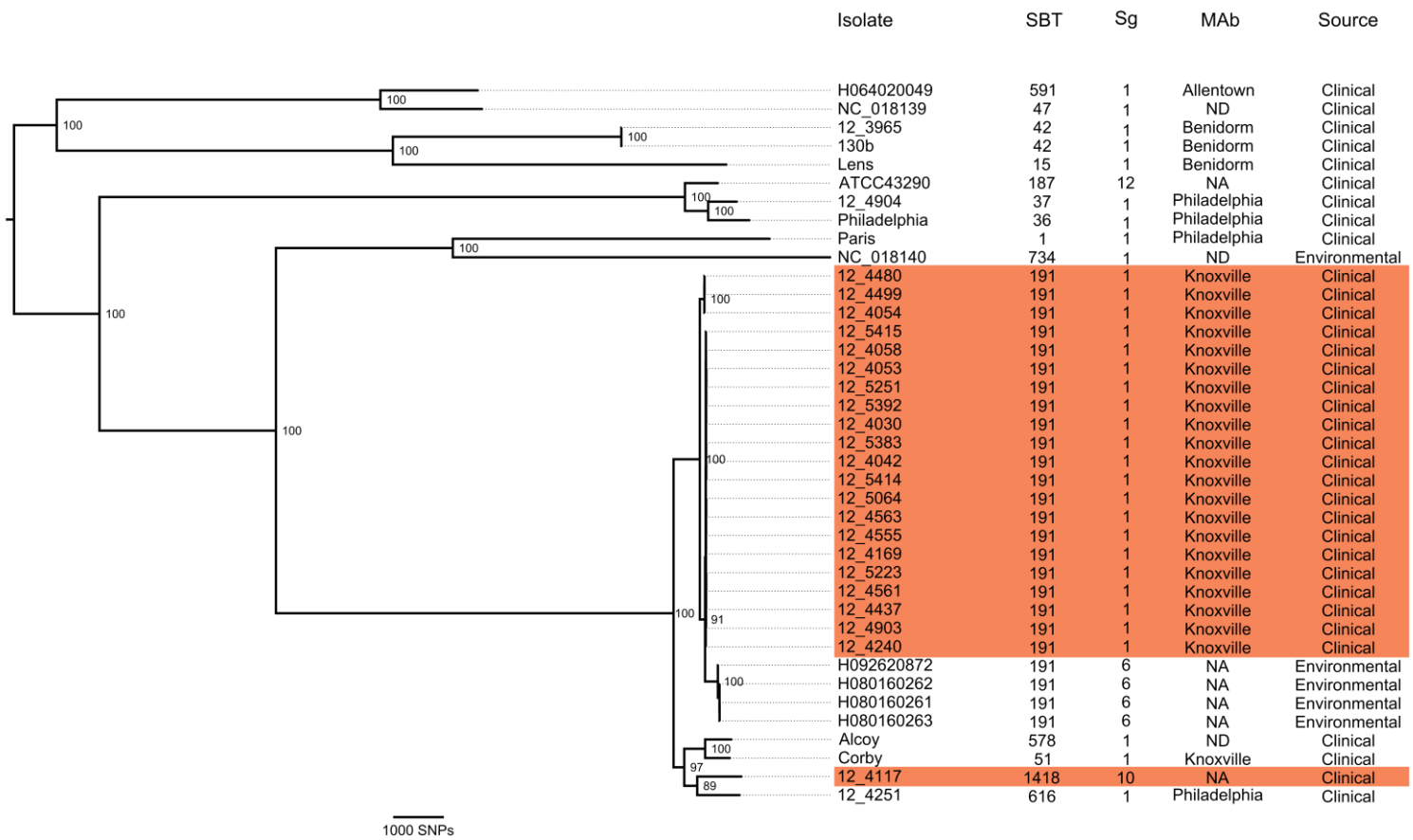


Figure 4.1 Maximum-likelihood phylogeny of *L. pneumophila*, labelled with results of traditional typing schemes. Outbreak isolates from Edinburgh are shaded in red. Tree is rooted on the branch leading to a *L. longbeachae* isolate used as an outgroup. ND: Not determined. NA: Not applicable.

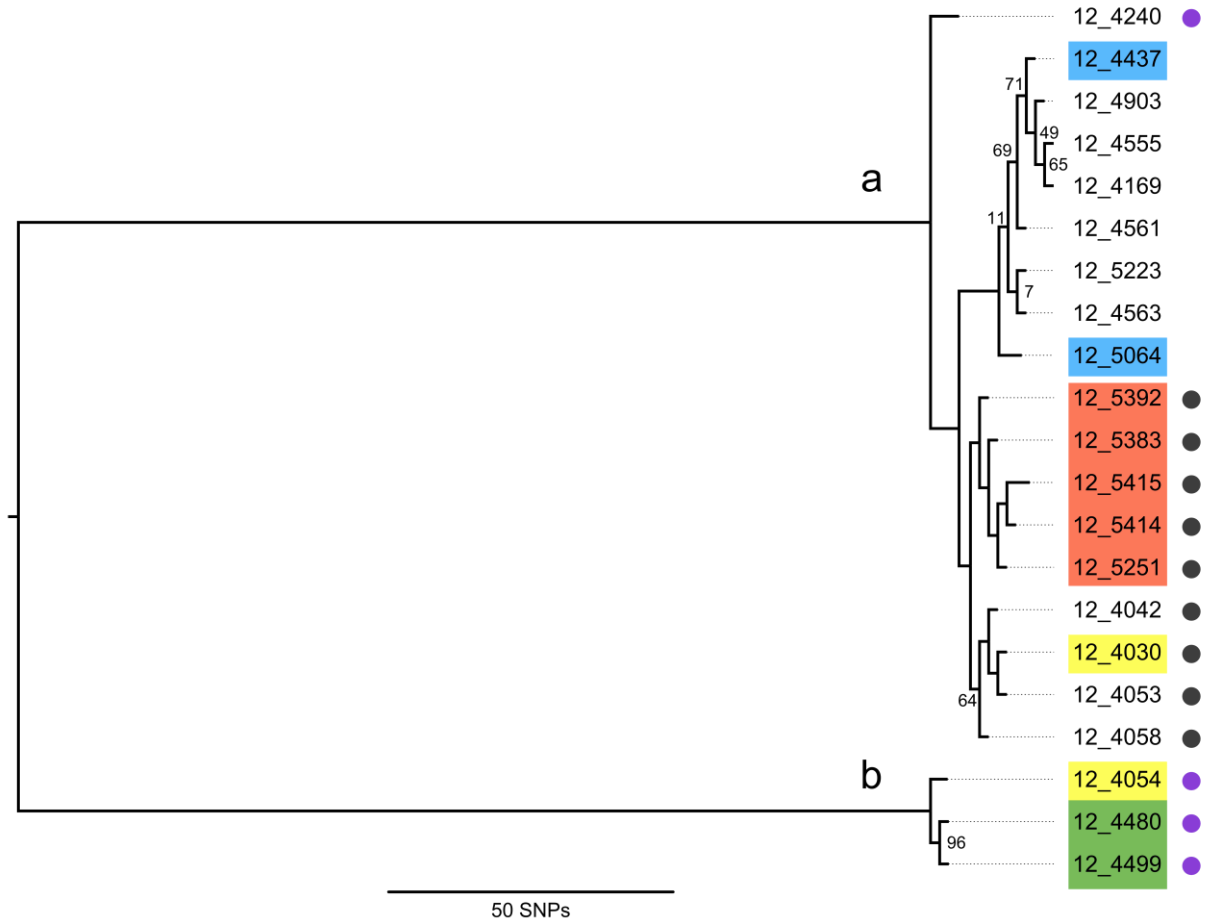


Figure 4.2 Maximum-likelihood phylogeny based on non-recombinant core genome of ST191 outbreak isolates. Two distinct clades are identifiable, labelled a and b. Colours correspond to patients for which there were multiple isolates: Yellow, patient 2; Green, patient 12; Blue, patient 13; Red, patient 1. Presence of an Lvh T4SS and a novel T4SS are denoted by black and purple shaded circles respectively.

4.4.2. Outbreak infections can be caused by multiple genetic subtypes of *L. pneumophila*

Multiple isolates were available for sequencing from patients 2, 12, 13 and 18 (Table 4.1 and Figure 4.2). Isolates 12_4030 and 12_4054 were obtained from patient 2, and belonged to clades A and B respectively (Figure 4.2), indicating either a heterogeneous infecting population from a single source, or exposure to 2 sources harbouring different populations of *L. pneumophila*. More limited heterogeneity was observed among the isolates from patients 13 and 18, indicative of within patient diversification from a single infecting strain, or acquisition of a closely related but heterogeneous infecting population.

The isolate 12_4117 was typed prior to whole genome sequencing and found to be Sg 1 ST191. However, the nucleotide sequence of the SBT loci from the whole genome sequence of the isolate was inconsistent with ST191, and typed the isolate as ST1418. To examine this discrepancy, multiple colonies from the original culture from the patient was reassessed by PCR and Sanger sequencing, which confirmed the presence of 2 distinct strains of *L. pneumophila* infecting this patient. This finding is suggestive of simultaneous infection of an individual patient with multiple STs of *L. pneumophila*.

Taken together the unexpected levels of diversity among isolates from this outbreak suggest that individuals were exposed to sources containing heterogeneous populations of *L. pneumophila*.

4.4.3. Examination of the accessory genome reveals variation in the complement of type 4 secretion systems among ST191 outbreak isolates

Variation in gene content was observed among the ST191 outbreak isolates with 2 distinct T4SS identified as being variably present in a subset of the outbreak isolates. A subset of 9 isolates from clade A (Figure 4.2), had 38 predicted CDSs with an average of 99.3% pairwise identity to a previously characterised T4SS that carries *lvh*, a locus involved in host cell entry, intracellular replication, and establishing infection at lower temperatures (Bandyopadhyay et al., 2007; Chien et al., 2004b; Ridenour et al., 2003). In addition, isolates belonging to clade B, and isolate 12_4240 from clade A (Figure 4.2), carried a previously uncharacterised T4SS with 45 predicted CDSs, including orthologs of *lvrA*, *lvrB*, *lvrC*, and *virB4* (Figure 4.3). The best scoring BLASTx results for the predicted *lvr* CDSs are to *Legionella drancourtii*, *Legionella longbeachae* and *L. pneumophila* but none show greater than 66% identity at the amino acid level.

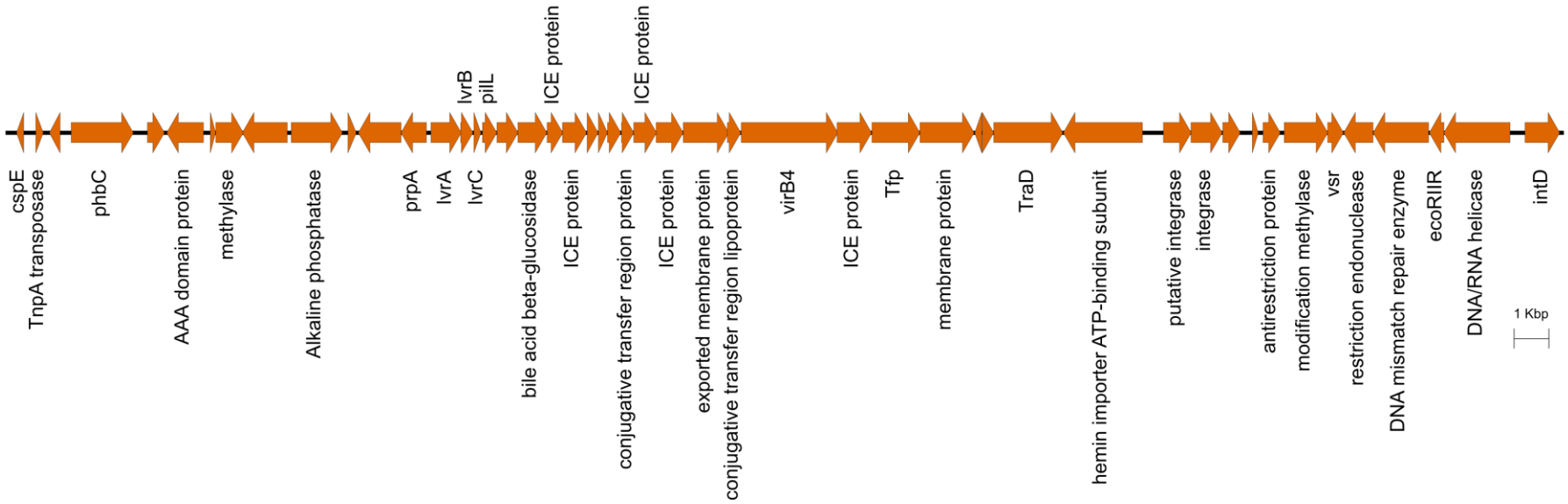


Figure 4.3 Schematic diagram of the novel T4SS present in isolates from clade B, and isolate 12_4240. Arrows represent predicted CDSs identified using the prokka annotation pipeline. Unlabelled CDSs are predicted hypothetical proteins. Figure produced using EasyFig.

4.4.4. Comparison of traditional typing methods with comparative genomic and phylogenetic analyses

The traditional typing methods of SBT, mAb typing and serotyping were compared with comparative genomic and phylogenetic analyses in order to determine the level of discriminatory power of each method (Figure 4.1)

21/22 of the outbreak isolates were typed as ST191, mAb subtype Knoxville and Sg 1. The remaining outbreak isolate was typed as ST1418 and Sg 10. Non-outbreak isolates from Scotland could be differentiated from the Edinburgh outbreak isolates based on differences in the SBT profiles and mAb subtypes. The ST191 environmental isolates were typed as Sg 6, and so were distinguishable from the Edinburgh outbreak based on serogroup.

However, SBT, mAb typing and serotyping failed to resolve the ST191 Edinburgh outbreak isolates into 2 distinct clades, and were unable to determine the presence or absence of accessory genomic elements among the outbreak isolates.

4.5. Discussion

Whole-genome sequencing of bacterial pathogens is increasingly being applied to the investigation of infectious disease outbreaks. The sequencing of *L. pneumophila* isolates obtained over the time course of a LD outbreak in Edinburgh in the current study added valuable new insights into the dynamics of the outbreak.

Outbreak isolates were readily distinguished from non-outbreak isolates on the basis of their position in the phylogenetic tree. 95% (21/22) of outbreak isolates were Sg 1, ST191, mAb Knoxville. Sg 1 is internationally disseminated and has previously been reported to be associated with human infections. For example, a study in France reported that 28.2% of environmental *Legionella* spp. isolates were *L. pneumophila* Sg 1, compared to 95.4% of clinical isolates (Doleans et al., 2004). However, ST191 is not commonly associated with human infection, but has been detected in environmental samples from the UK, Germany, the Netherlands, Poland, and Russia (PHE Legionella Database). The wide dissemination of ST191 in the environment, and its apparent pathogenicity highlights its potential for future outbreaks.

The current study clearly illustrates the potential of genomic epidemiology to add significant power to the investigation of infectious disease outbreaks by providing a higher resolution view of the *L. pneumophila* population than is possible through the use of traditional techniques such as SBT, serotyping and monoclonal antibody typing. The levels of genetic variation within the isolates from 4 of the patients raise questions as to how these levels of diversity arose. It has been demonstrated that a cloud of variation exists within the infecting populations of patients with bacterial infections (Harris et al., 2013; Young et al., 2012), and that this heterogeneity may arise through either within-host pathogen evolution, or from multiple pathogen acquisition events (Bryant et al., 2013; Lieberman et al., 2011). This is an unlikely scenario for LD outbreaks given the lack of person-to-person transmission for *L. pneumophila,* and the timescale between exposure to and isolation of the pathogen

during the Edinburgh outbreak. Alternatively, the initial exposure may be to a heterogeneous *L. pneumophila* population. Multiple genotypes of *L. pneumophila*, can persist in cooling towers for periods of at least 5 years, allowing for high levels of genetic diversity within a single environmental source (Sanchez et al., 2008). In addition, the same study demonstrated the presence of *L. pneumophila* with indistinguishable PFGE patterns in multiple cooling towers within a 1 km radius, suggesting that the possibility of cross-contamination of water reservoirs needs to be accounted for during outbreak investigations (Sanchez et al., 2008).

Intra- and inter-species horizontal gene transfer of is common within *L. pneumophila*, and variations in T4SS content between strains have been reported previously (Bandyopadhyay et al., 2007; D'Auria et al., 2010; Schroeder et al., 2010). However, differences in T4SS content have not previously been identified among strains from a single outbreak. Within the ST191 outbreak isolates, variation in content of 2 different T4SS was identified. The phenotypic effect of this variation remains to be determined, but it is possible that the strains may display differences in pathogenic potential. T4SS can deliver effector proteins into host cells and are involved in host cell entry, intracellular replication, modulation of host immunity, and establishing infection at lower temperatures (Bandyopadhyay et al., 2007; Fronzes et al., 2009; Lomma et al., 2009; Ridenour et al., 2003).

The levels of variation identified within the Edinburgh outbreak raise important issues relating to outbreak detection, and also disease progression within the patient. The effect on the patient of co-infection with multiple strains of *L. pneumophila* is

not currently known and further investigation is required in order to determine the prevalence of co-infection within this outbreak, and whether co-infection can be used as a predictive indicator of disease progression. Attenuation of the pathogenic potential of T4SS mutants in comparison to wild-type *L. pneumophila* strains have been observed using the larvae of the wax moth *Galleria mellonella* as an infection model (Harding et al., 2012). This infection model may be a suitable system in which to test the virulence phenotype of the isolates obtained for this study.

While the methods presented here have clear advantages over traditional typing methods, there are a number of limitations with the current study. The full level of diversity within each patient has not been sampled here, but the methods described offer the opportunity to do so and could shed further light on the potential source of the outbreak.

The use of the whole genome sequencing approach may be limited in the investigation of LD outbreaks. Clinical laboratories do not routinely culture specimens due to the time and expertise required to isolate *L. pneumophila* from sputum, but this may change if the information generated can have a meaningful impact on transmission control or patient treatment. Isolates were only successfully cultured from 18/91 (19.8%) of patients with a confirmed or suspected case of LD, raising the likelihood that a level of *L. pneumophila* variation within the outbreak remains to be captured. In addition, the use of whole genome sequencing to identify the source of an outbreak is currently dependent on a viable culture from a suspected source. In the current study, *L. pneumophila* was not successfully cultured from

suspected sources, possibly due to water samples being collected after *Legionella* control measures had been implemented. It would perhaps be useful to routinely sample *L. pneumophila* reservoirs in order to build up a database of Legionella diversity, to which sequences of isolates from any subsequent outbreaks can be compared.

The limitations of *L. pneumophila* culture could be overcome by culture-free sequencing methodologies. It has recently been demonstrated that pathogen genome sequences can be derived directly from clinical samples, without the need for culturing (Loman et al., 2013; Seth-Smith et al., 2013). Application of culture-free sequencing has the potential to revolutionise the study of difficult to culture pathogens, and could be used to obtain whole genome sequences of *L. pneumophila* directly from a patient's sputum or from environmental water sources.

4.6. Conclusions

Whole genome sequencing presents the opportunity to revolutionise the investigation of bacterial epidemics. Being able to identify potential sources of infection, and transmission pathways allows for the design of effective control measures that can halt the progress of an outbreak, reducing disease burden. The current study suggests that the diversity within this *L. pneumophila* outbreak may preclude the use of WGS for the definitive identification of a source reservoir, and could be of limited benefit for resolving legal issues relating to outbreaks. Characterising the levels of diversity among epidemic strains, and comparisons with non-epidemic strains may lead to the identification of molecular correlates of outbreak success. Routine

sequencing of environmental water reservoirs could lead to shifts in bacterial populations being detected prior to an epidemic, allowing for pre-emptive treatment of reservoirs.

5. General discussion

5.1. Revolutionising infectious disease epidemiology through WGS

The emergence of bacterial pathogens, together with the re-emergence of existing clones with altered resistance or virulence phenotypes pose a serious threat to the health of the global population. In order to develop novel therapeutic agents and effective infection control measures, it is essential to understand the evolution and dynamics of pathogenic bacterial clones. In particular, the impact of anthropogenic practices such as antimicrobial prescription and vaccination programmes on the emergence of clones is not well understood. Throughout the last century, the improvement in the treatment of bacterial disease was driven by increased levels of nutrition and sanitation, and through the development of effective vaccines and antimicrobials. Currently, standards of nutrition and sanitation in the developed world are unlikely to improve much further, necessitating the development of novel methods for pathogen control. The current work provides examples of the application of WGS of bacterial populations to further the understanding of bacterial evolution on different scales of space and time, highlighting the advances possible from the level of a single patient, to the study of pandemic clones affecting thousands of people.

5.2. Bacterial WGS in the clinic

The study of bacterial evolution during infection has been limited by the resolution afforded by traditional molecular typing methods, which are unable to identify small-scale genetic changes that may have profound effects on host-pathogen interactions. In the past, monoclonal infections have been assumed to involve a single infecting strain and the high-resolution technologies have not existed to characterise the

variation levels present within an infection. In reality, there likely exists a 'cloud' of diversity within an infecting bacterial population that must be considered when characterising host-pathogen interactions (Bryant et al., 2013; Lieberman et al., 2011). Information derived from the analysis of WGS has numerous potential implications in clinical settings. WGS vastly increases the resolution of genetic variation that can be determined in comparison to traditional molecular typing methods (Lieberman et al., 2011; McAdam et al., 2011; Mwangi et al., 2007).

An understanding of the heterogeneity of the infecting population may ultimately inform on disease progression and outcome. It has previously been demonstrated that microorganisms can evolve differently under the selective pressures found in different hosts (Zdziarski et al., 2010) and this has implications for the development of personalised genomic medicine, and tailored therapies. In contrast, convergent evolution of pathogens during long term infection of independent hosts has been described (Lieberman et al., 2011) indicating potential new targets for therapeutics. The current work demonstrates the variation that is present during chronic *S. aureus* infection of a cystic fibrosis patient (McAdam et al., 2011). In particular, differences in resistance and virulence phenotype were observed, highlighting the need to think of a bacterial infection in terms of a diverse population, rather than as a monoclonal organism. Within the patient, the SigB locus appeared to be under selection in the isolates from the study patient, highlighting a potential target for antimicrobial therapy. An integrated systems biology approach of linking polymorphisms to virulence would aid in the prediction of disease progression (Priest et al., 2012),

phenotypic testing in conjunction with genotyping could provide a framework for developing predictive models.

In addition, the ability to predict the antibiotic resistance profile of an isolate from its genome sequence has been demonstrated with high levels of accuracy (Holden et al., 2013; Köser et al., 2012) and the full repertoire of toxins carried by an isolate may be identified, providing information relevant to pathogenic potential and disease outcome (Köser et al., 2012). In the current study, a fusidic acid resistant phenotype emerged during the course of infection, and the specific polymorphism conferring in resistance was identified in the *fusA* gene (McAdam et al., 2011). Ultimately, bacterial WGS may replace traditional methods of determining resistance, due to the high throughput capacity, and decreasing cost (Loman et al., 2012b; Priest et al., 2012).

5.3. WGS of bacteria as a tool for outbreak detection and molecular epidemiology

Comparative genomic and phylogenetic analyses of bacterial populations have resulted in the identification of transmission pathways within hospitals, countries, and across continents (Harris et al., 2010; Holden et al., 2013; Lewis et al., 2010; McAdam et al., 2011). Network modelling approaches have shown that the infection rate of a given hospital is positively correlated with its level of connectedness to other hospitals in the network (Donker et al., 2010, 2012). This is consistent with the findings of the current work, which identified large population centres within the UK as hubs of transmission for the hospital restricted EMRSA-16 lineage (McAdam et

al., 2012). Patient referrals may be the reason for the observed pattern of transmission, but the development of methodologies integrating phylogenetic and epidemiological models could facilitate formal evaluation of risk factors. Methods integrating predictive epidemiological models with empirical data from WGS can simultaneously reconstruct spatiotemporal history and evaluate the impact of risk factors for pathogen spread. For example, a phylogeographic model for the seasonal emergence and spread of H3N2 influenza that incorporated sequence data, isolate location, population size, and airline passenger flux data identified passenger flux as a major contributor to global influenza transmission (Lemey et al., 2012). The spatial spread of H1N1 and H1N2 in US swine populations has been shown to follow long distance movements of swine populations from southeast and south-central USA to the Midwest (Nelson et al., 2011). By integrating sequence data and epidemiological information, risk factors for pathogen spread can be objectively evaluated using empirical evidence, which could ultimately be used to influence the rational design of control programs.

Many pathogens have emerged as a direct result of human practices such as industrialisation and globalisation, as highlighted by the rapid spread of dengue virus from regions of Southeast Asia and South America since the 1960s. A combination of demographic factors has resulted in conditions favouring viral transmission by the mosquito vector *Aedes aegypti*. Rapid human population growth, coupled with increased rural to urban migration has resulted large, densely populated areas, while increases in solid waste has created habitats within urban centres capable of supporting the vector. Subsequently, global geographic expansion of the mosquito

vector has been driven through international trade routes (Gubler, 2011). Similarly, Legionnaire's disease is a product of the manmade environment. Increased industrialisation has seen the widespread construction of cooling towers that transfer excess heat generated by machinery to the atmosphere, and offer a habitat for *Legionella* species in the midst of large human populations. Furthermore, measures designed to increase human comfort such as air conditioning units offer an ideal habitat for growth of *Legionella* species, and a ready means of dispersal to large numbers of hosts (Ward et al., 2010). Previously, sources of epidemic *L. pneumophila* have been identified through the use of traditional typing methods (Hugosson et al., 2007; White et al., 2013). However, the high-resolution afforded by WGS of isolates from the 2012 Edinburgh outbreak of Legionnaire's disease in the current work revealed unprecedented levels of diversity, suggesting diversification of *L. pneumophila* over a time period greatly in excess of the timescale of the outbreak. This has important implications for outbreak investigation, it may well prove impossible to definitively identify a single location as the likely source of a *L. pneumophila* epidemic, unless the full diversity of environmental *L. pneumophila* can be sequenced. Even if the full diversity of *L. pneumophila* can be characterised, the possibility of cross-contamination of water reservoirs cannot be excluded. A study of cooling towers in Spain showed the presence of *L. pneumophila* with indistinguishable PFGE patterns in multiple sites within a 1 km radius (Sanchez et al., 2008). Taken together, these observations suggest that WGS may be unable to determine the source of *L. pneumophila* epidemics, and could be of limited benefit for resolving legal issues relating to outbreaks.

5.4. Using bacterial WGS to study pandemic clone emergence and adaptation

The impact of globalisation on the dynamics of infectious disease is incompletely understood. Previously disparate populations are becoming increasingly interconnected as a result of the easy availability of global travel. Prior to the widespread use of air travel, an individual's social network was relatively self-contained with a predictable level of organisation. However, increased range and frequency of travel increases the complexity of individual social networks by introducing an element of randomness, resulting in 'small-world' networks, where 2 individuals are linked by a relatively small number of mutual connections (Watts et al., 1998). Modelling of infectious disease transmission throughout these networks has identified a decrease in the level of infectivity that is required to infect 50% of a susceptible population, and a decrease in the time taken to do so (Watts et al., 1998). The current work describes the global distribution of 3 lineages of *S. aureus* CC30, and identifies an independent origin for each pandemic clone, highlighting the potential of successful clones to rapidly disseminate and cause an important clinical burden in many regions (McAdam et al., 2012).

Advancements in medical interventions designed to reduce disease burden can inadvertently promote the increased prevalence of pathogenic bacterial clones. Vaccines that are ineffective against the full range of diversity present in the population can select for the population expansion of resistant strains that fill the niche vacated by susceptible lineages (Croucher et al., 2011; Golubchik et al., 2012).

In the current work, Bayesian coalescent analyses were employed to determine the dates of emergence of important pandemic lineages of *S. aureus* CC30, and determine whether they coincide with changes in medical practice. Specifically, the emergence of the penicillin-resistant phage-type 80/81 clone was predicted to be commensurate with the introduction of penicillin as an antimicrobial therapy (McAdam et al., 2012).

The role of horizontal gene transfer has been demonstrated in niche adaptation (Guinane et al., 2010; Lowder et al., 2009) and in the acquisition of antibiotic resistance and virulence traits (Baba et al., 2002; Chang et al., 2003; Fitzgerald et al., 2001; Holden et al., 2004; Mutreja et al., 2011; Rohde et al., 2011). The current work has identified molecular correlates of pandemic lineage emergence and niche restriction and was dependent on a large genome sequence set for comparisons (McAdam et al., 2012). Specifically, the adaptation of EMRSA-16 to the hospital environment can be linked to the acquisition of the type II SCC*mec* element that confers multidrug resistance, while a number of mutations that could result in a low virulence phenotype were also described (McAdam et al., 2012).

The full potential of WGS of bacterial populations depends on the ability to place strains in their appropriate temporal and spatial context in order to identify the processes that led to their selection and emergence. An understanding of the core genome of a bacterial species is required in order to determine molecular correlates that may lead to adaptation to novel niches.

During the 2011 European O104:H4 *E. coli* outbreak, the success in identifying the cause of increased virulence in the O104:H4 *E. coli* strains was dependent on researchers having a solid understanding of related genomes and having data from other sequenced genomes available for comparisons (Rohde et al., 2011). In order for rapid comparisons of new sequences to be achievable, there needs to be consensus across the research community with regards to standards of sequencing, genome assembly and data availability.

5.5. The future of bacterial WGS

A major drawback to the WGS approach using current widely available sequencing technologies is that sequence assemblies are typically incomplete. Repeat regions of pathogen genomes preclude a single contiguous assembly of short read data, and additional uncertainty can be introduced in a number of ways. Inaccurate nucleotide base calls may be introduced through poor filtering of raw data, mapping of sequence reads to multiple regions of the genome, and from errors inherent in the sequencing technology used. Additionally, identification of novel strain specific elements is not possible using a mapping approach, and requires genomes to be assembled *de novo*, with subsequent annotation of contigs. Single-molecule sequencing technologies have the capability to sequence bacterial genomes on a single contig and to a finished-grade quality (Koren et al., 2013). However, the costs are currently prohibitively expensive for routine use.

5.6. Conclusions

In conclusion, this work demonstrates the utility of applying WGS to whole bacterial populations and offers new insights into pathogen evolution within host, and within the larger host population. Molecular correlates of pathogen emergence have been identified and linked to changes in human activity. It has important implications for the design of control measures, therapeutic and public health interventions should be rationally designed so as to prevent the emergence of lineages with elevated levels of antibiotic resistance or virulence. The ultimate use of WGS of bacterial pathogens would be to identify molecular correlates of disease progression at the single host level, and with the emergence of virulent clones at the population level.

6. References

AIRES DE SOUSA, M., CRISOSTOMO, M. I., SANCHES, I. S., WU, J. S., FUZHONG, J., TOMASZ, A. & DE LENCASTRE, H. 2003. Frequent recovery of a single clonal type of multidrug-resistant *Staphylococcus aureus* from patients in two hospitals in Taiwan and China. *J Clin Microbiol,* 41**,** 159-63.

AIRES DE SOUSA, M., MIRAGAIA, M., SANCHES, I. S., AVILA, S., ADAMSON, I., CASAGRANDE, S. T., BRANDILEONE, M. C., PALACIO, R., DELL'ACQUA, L., HORTAL, M., CAMOU, T., ROSSI, A., VELAZQUEZ-MEZA, M. E., ECHANIZ-AVILES, G., SOLORZANO-SANTOS, F., HEITMANN, I. & DE LENCASTRE, H. 2001. Three-year assessment of methicillin-resistant *Staphylococcus aureus* clones in Latin America from 1996 to 1998. *J Clin Microbiol,* 39**,** 2197-205.

AIRES DE SOUSA, M., SANCHES, I. S., FERRO, M. L., VAZ, M. J., SARAIVA, Z., TENDEIRO, T., SERRA, J. & DE LENCASTRE, H. 1998. Intercontinental spread of a multidrug-resistant methicillin-resistant *Staphylococcus aureus* clone. *J Clin Microbiol,* 36**,** 2590-6.

ALESANA-SLATER, J., RITCHIE, S. R., HEFFERNAN, H., CAMP, T., RICHARDSON, A., HERBISON, P. & NORRIS, P. 2011. Methicillin-resistant *Staphylococcus aureus*, Samoa, 2007-2008. *Emerg Infect Dis*, 17**,** 1023-9.

ALIKHAN, N. F., PETTY, N. K., BEN ZAKOUR, N. L. & BEATSON, S. A. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics,* 12**,** 402.

ALP, E., KLAASSEN, C. H., DOGANAY, M., ALTOPARLAK, U., AYDIN, K., ENGIN, A., KUZUCU, C., OZAKIN, C., OZINEL, M. A., TURHAN, O. & VOSS, A. 2009. MRSA genotypes in Turkey: persistence over 10 years of a single clone of ST239. *J Infect,* 58**,** 433-8.

AMARO, F., GILBERT, J. A., OWENS, S., TRIMBLE, W. & SHUMAN, H. A. 2012. Whole-genome sequence of the human pathogen *Legionella pneumophila* serogroup 12 strain 570-CO-H. *J Bacteriol,* 194**,** 1613-4.

ANGIUOLI, S. V., DUNNING HOTOPP, J. C., SALZBERG, S. L. & TETTELIN, H. 2011. Improving pan-genome annotation using whole genome multiple alignment. *BMC Bioinformatics,* 12**,** 272.

ANGIUOLI, S. V. & SALZBERG, S. L. 2011. Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics,* 27**,** 334-42.

ARONESTY, E. 2011. ea-utils. code.google.com/p/ea-utils/

AYLIFFE, G. A. 1997. The progressive intercontinental spread of methicillin-resistant *Staphylococcus aureus*. *Clin Infect Dis,* 24 Suppl 1**,** S74-9.

BABA, T., TAKEUCHI, F., KURODA, M., YUZAWA, H., AOKI, K., OGUCHI, A., NAGAI, Y., IWAMA, N., ASANO, K., NAIMI, T., KURODA, H., CUI, L., YAMAMOTO, K. & HIRAMATSU, K. 2002. Genome and virulence determinants of high virulence community-acquired MRSA. *Lancet,* 359**,** 1819-27.

BANDYOPADHYAY, P., LIU, S., GABBAI, C. B., VENITELLI, Z. & STEINMAN, H. M. 2007. Environmental mimics and the Lvh type IVA secretion system contribute to virulence-related phenotypes of *Legionella pneumophila*. *Infect Immun,* 75**,** 723-35.

BANERJEE, R., GRETES, M., HARLEM, C., BASUINO, L. & CHAMBERS, H. F. 2010. A *mecA*-negative strain of methicillin-resistant *Staphylococcus aureus* with high-level beta-lactam resistance contains mutations in three genes. *Antimicrob Agents Chemother,* 54**,** 4900-2.

BARNES, S. L., HARRIS, A. D., GOLDEN, B. L., WASIL, E. A. & FURUNO, J. P. 2011. Contribution of interfacility patient movement to overall methicillin-resistant *Staphylococcus aureus* prevalence levels. *Infect Control Hosp Epidemiol,* 32**,** 1073-8.

BEGIER, E. M., FRENETTE, K., BARRETT, N. L., MSHAR, P., PETIT, S., BOXRUD, D. J., WATKINS-COLWELL, K., WHEELER, S., CEBELINSKI, E. A., GLENNEN, A., NGUYEN, D., HADLER, J. L. & CONNECTICUT BIOTERRORISM FIELD EPIDEMIOLOGY RESPONSE TEAM. 2004. A high-morbidity outbreak of methicillin-resistant *Staphylococcus aureus* among players on a college football team, facilitated by cosmetic body shaving and turf burns. *Clin Infect Dis,* 39**,** 1446-53.

BENIN, A. L., BENSON, R. F. & BESSER, R. E. 2002. Trends in legionnaires disease, 1980-1998: declining mortality and new patterns of diagnosis. *Clin Infect Dis,* 35**,** 1039-46.

BERES, S. B., CARROLL, R. K., SHEA, P. R., SITKIEWICZ, I., MARTINEZ-GUTIERREZ, J. C., LOW, D. E., MCGEER, A., WILLEY, B. M., GREEN, K., TYRRELL, G. J., GOLDMAN, T. D., FELDGARDEN, M., BIRREN, B. W., FOFANOV, Y., BOOS, J., WHEATON, W. D., HONISCH, C. & MUSSER, J. M. 2010. Molecular complexity of successive bacterial epidemics deconvoluted by comparative pathogenomics. *Proc Natl Acad Sci U S A,* 107**,** 4371-6.

BERGDOLL, M. S., CRASS, B. A., REISER, R. F., ROBBINS, R. N. & DAVIS, J. P. 1981. A new staphylococcal enterotoxin, enterotoxin F, associated with toxic-shock-syndrome *Staphylococcus aureus* isolates. *Lancet,* 1**,** 1017-21.

BESIER, S., LUDWIG, A., BRADE, V. & WICHELHAUS, T. A. 2003. Molecular analysis of fusidic acid resistance in *Staphylococcus aureus*. *Mol Microbiol,*

47**,** 463-9.

BESIER, S., SMACZNY, C., VON MALLINCKRODT, C., KRAHL, A., ACKERMANN, H., BRADE, V. & WICHELHAUS, T. A. 2007. Prevalence and clinical significance of *Staphylococcus aureus* small-colony variants in cystic fibrosis lung disease. *J Clin Microbiol,* 45**,** 168-72.

BESIER, S., ZANDER, J., KAHL, B. C., KRAICZY, P., BRADE, V. & WICHELHAUS, T. A. 2008. The thymidine-dependent small-colony-variant phenotype is associated with hypermutability and antibiotic resistance in clinical *Staphylococcus aureus* isolates. *Antimicrob Agents Chemother,* 52**,** 2183-9.

BLAIR, J. E. & CARR, M. 1960. Distribution of Phage Groups of *Staphylococcus aureus* in the Years 1927 through 1947. *Science,* 132**,** 1247-8.

BOISSY, R., AHMED, A., JANTO, B., EARL, J., HALL, B. G., HOGG, J. S., PUSCH, G. D., HILLER, L. N., POWELL, E., HAYES, J., YU, S., KATHJU, S., STOODLEY, P., POST, J. C., EHRLICH, G. D. & HU, F. Z. 2011. Comparative supragenomic analyses among the pathogens *Staphylococcus aureus*, *Streptococcus pneumoniae,* and *Haemophilus influenzae* using a modification of the finite supragenome model. *BMC Genomics,* 12**,** 187.

BOYLE-VAVRA, S. & DAUM, R. S. 2007. Community-acquired methicillin-resistant *Staphylococcus aureus*: the role of Panton-Valentine leukocidin. *Lab Invest,* 87**,** 3-9.

BRAN, J. L., LEVISON, M. E. & KAYE, D. 1972. Survey for methicillin-resistant staphylococci. *Antimicrob Agents Chemother,* 1**,** 235-6.

BRANGER, C., GARDYE, C. & LAMBERT-ZECHOVSKY, N. 1996. Persistence

of *Staphylococcus aureus* strains among cystic fibrosis patients over extended periods of time. *J Med Microbiol,* 45**,** 294-301.

BROWN, C. M., NUORTI, P. J., BREIMAN, R. F., HATHCOCK, A. L., FIELDS, B. S., LIPMAN, H. B., LLEWELLYN, G. C., HOFMANN, J. & CETRON, M. 1999. A community outbreak of Legionnaires' disease linked to hospital cooling towers: an epidemiological method to calculate dose of exposure. *Int J Epidemiol,* 28**,** 353-9.

BRUGGEMANN, H., CAZALET, C. & BUCHRIESER, C. 2006. Adaptation of *Legionella pneumophila* to the host environment: role of protein secretion, effectors and eukaryotic-like proteins. *Curr Opin Microbiol,* 9**,** 86-94.

BRYANT, J. M., GROGONO, D. M., GREAVES, D., FOWERAKER, J., RODDICK, I., INNS, T., REACHER, M., HAWORTH, C. S., CURRAN, M. D., HARRIS, S. R., PEACOCK, S. J., PARKHILL, J. & FLOTO, R. A. 2013. Whole-genome sequencing to identify transmission of *Mycobacterium abscessus* between patients with cystic fibrosis: a retrospective cohort study. *Lancet,* 381**,** 1551-1560.

BURGDORFER, W., BARBOUR, A. G., HAYES, S. F., BENACH, J. L., GRUNWALDT, E. & DAVIS, J. P. 1982. Lyme disease-a tick-borne spirochetosis? *Science,* 216**,** 1317-9.

CARLETON, H. A., DIEP, B. A., CHARLEBOIS, E. D., SENSABAUGH, G. F. & PERDREAU-REMINGTON, F. 2004. Community-adapted methicillin-resistant *Staphylococcus aureus* (MRSA): population dynamics of an expanding community reservoir of MRSA. *J Infect Dis,* 190**,** 1730-8.

CAZALET, C., RUSNIOK, C., BRUGGEMANN, H., ZIDANE, N., MAGNIER, A., MA, L., TICHIT, M., JARRAUD, S., BOUCHIER, C., VANDENESCH, F., KUNST, F., ETIENNE, J., GLASER, P. & BUCHRIESER, C. 2004.

Evidence in the *Legionella pneumophila* genome for exploitation of host cell functions and high genome plasticity. *Nat Genet,* 36**,** 1165-73.

CDC 1999. Achievements in Public Health, 1900-1999: Control of Infectious Diseases. *MMWR. Morbidity and mortality weekly report,* 48**,** 621-629.

CDRW 1997. Epidemic methicillin resistant *Staphylococcus aureus*. *Commun Dis Rep CDR Wkly.,* 7**,** 191-191.

CHAMBERS, H. F. & DELEO, F. R. 2009. Waves of resistance: *Staphylococcus aureus* in the antibiotic era. *Nat Rev Microbiol,* 7**,** 629-41.

CHANG, S., SIEVERT, D. M., HAGEMAN, J. C., BOULTON, M. L., TENOVER, F. C., DOWNES, F. P., SHAH, S., RUDRIK, J. T., PUPP, G. R., BROWN, W. J., CARDO, D., FRIDKIN, S. K. & VANCOMYCIN-RESISTANT *STAPHYLOCOCCUS AUREUS* INVESTIGATIVE TEAM 2003. Infection with vancomycin-resistant *Staphylococcus aureus* containing the *vanA* resistance gene. *N Engl J Med,* 348**,** 1342-7.

CHEN, L., CHAVDA, K. D., SOLANKI, M., MEDIAVILLA, J. R., MATHEMA, B., SCHLIEVERT, P. M. & KREISWIRTH, B. N. 2013. Genetic variation among Panton-Valentine leukocidin-encoding bacteriophages in *Staphylococcus aureus* clonal complex 30 strains. *J Clin Microbiol,* 51**,** 914-9.

CHIEN, M., MOROZOVA, I., SHI, S., SHENG, H., CHEN, J., GOMEZ, S. M., ASAMANI, G., HILL, K., NUARA, J., FEDER, M., RINEER, J., GREENBERG, J. J., STESHENKO, V., PARK, S. H., ZHAO, B., TEPLITSKAYA, E., EDWARDS, J. R., PAMPOU, S., GEORGHIOU, A., CHOU, I. C., IANNUCCILLI, W., ULZ, M. E., KIM, D. H., GERINGER-SAMETH, A., GOLDSBERRY, C., MOROZOV, P., FISCHER, S. G., SEGAL, G., QU, X., RZHETSKY, A., ZHANG, P., CAYANIS, E., DE

JONG, P. J., JU, J., KALACHIKOV, S., SHUMAN, H. A. & RUSSO, J. J. 2004. The genomic sequence of the accidental pathogen *Legionella pneumophila*. *Science,* 305**,** 1966-8.

COLEMAN, D. C., SULLIVAN, D. J., RUSSELL, R. J., ARBUTHNOTT, J. P., CAREY, B. F. & POMEROY, H. M. 1989. *Staphylococcus aureus* bacteriophages mediating the simultaneous lysogenic conversion of beta-lysin, staphylokinase and enterotoxin A: molecular mechanism of triple conversion. *J Gen Microbiol,* 135**,** 1679-97.

COLLIGNON, P., GOSBELL, I., VICKERY, A., NIMMO, G., STYLIANOPOULOS, T. & GOTTLIEB, T. 1998. Community-acquired meticillin-resistant *Staphylococcus aureus* in Australia. Australian Group on Antimicrobial Resistance. *Lancet,* 352**,** 145-6.

COLLINS, J., RUDKIN, J., RECKER, M., POZZI, C., O'GARA, J. P. & MASSEY, R. C. 2010. Offsetting virulence and antibiotic resistance costs by MRSA. *ISME J,* 4**,** 577-84.

CONNOR, R. I., SHERIDAN, K. E., CERADINI, D., CHOE, S. & LANDAU, N. R. 1997. Change in coreceptor use correlates with disease progression in HIV-1--infected individuals. *J Exp Med,* 185**,** 621-8.

CONZA, L., CASATI, S., LIMONI, C. & GAIA, V. 2013. Meteorological factors and risk of community-acquired Legionnaires' disease in Switzerland: an epidemiological study. *BMJ Open,* 3.

COOKSON, B. D., ROBINSON, D. A., MONK, A. B., MURCHAN, S., DEPLANO, A., DE RYCK, R., STRUELENS, M. J., SCHEEL, C., FUSSING, V., SALMENLINNA, S., VUOPIO-VARKILA, J., CUNY, C., WITTE, W., TASSIOS, P. T., LEGAKIS, N. J., VAN LEEUWEN, W., VAN BELKUM, A., VINDEL, A., GARAIZAR, J., HAEGGMAN, S., OLSSON-

LILJEQUIST, B., RANSJO, U., MULLER-PREMRU, M., HRYNIEWICZ, W., ROSSNEY, A., O'CONNELL, B., SHORT, B. D., THOMAS, J., O'HANLON, S. & ENRIGHT, M. C. 2007. Evaluation of molecular typing methods in characterizing a European collection of epidemic methicillin-resistant *Staphylococcus aureus* strains: the HARMONY collection. *J Clin Microbiol,* 45**,** 1830-7.

COUTO, I., COSTA, S. S., VIVEIROS, M., MARTINS, M. & AMARAL, L. 2008. Efflux-mediated response of *Staphylococcus aureus* exposed to ethidium bromide. *J Antimicrob Chemother,* 62**,** 504-13.

COX, R. A., CONQUEST, C., MALLAGHAN, C. & MARPLES, R. R. 1995. A major outbreak of methicillin-resistant *Staphylococcus aureus* caused by a new phage-type (EMRSA-16). *J Hosp Infect,* 29**,** 87-106.

CROUCHER, N. J., HARRIS, S. R., FRASER, C., QUAIL, M. A., BURTON, J., VAN DER LINDEN, M., MCGEE, L., VON GOTTBERG, A., SONG, J. H., KO, K. S., PICHON, B., BAKER, S., PARRY, C. M., LAMBERTSEN, L. M., SHAHINAS, D., PILLAI, D. R., MITCHELL, T. J., DOUGAN, G., TOMASZ, A., KLUGMAN, K. P., PARKHILL, J., HANAGE, W. P. & BENTLEY, S. D. 2011. Rapid pneumococcal evolution in response to clinical interventions. *Science,* 331**,** 430-4.

D'AURIA, G., JIMENEZ-HERNANDEZ, N., PERIS-BONDIA, F., MOYA, A. & LATORRE, A. 2010. *Legionella pneumophila* pangenome reveals strain-specific virulence factors. *BMC Genomics,* 11**,** 181.

DARLING, A. E., MAU, B. & PERNA, N. T. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One,* 5**,** e11147.

DAVIES, J. & DAVIES, D. 2010. Origins and evolution of antibiotic resistance.

Microbiol Mol Biol Rev, 74**,** 417-33.

DAVIS, J. P., CHESNEY, P. J., WAND, P. J. & LAVENTURE, M. 1980. Toxic-shock syndrome: epidemiologic features, recurrence, risk factors, and prevention. *N Engl J Med,* 303**,** 1429-35.

DAVIS, K. A., STEWART, J. J., CROUCH, H. K., FLOREZ, C. E. & HOSPENTHAL, D. R. 2004. Methicillin-resistant *Staphylococcus aureus* (MRSA) nares colonization at hospital admission and its effect on subsequent MRSA infection. *Clin Infect Dis,* 39**,** 776-82.

DELEO, F. R., KENNEDY, A. D., CHEN, L., BUBECK WARDENBURG, J., KOBAYASHI, S. D., MATHEMA, B., BRAUGHTON, K. R., WHITNEY, A. R., VILLARUZ, A. E., MARTENS, C. A., PORCELLA, S. F., MCGAVIN, M. J., OTTO, M., MUSSER, J. M. & KREISWIRTH, B. N. 2011. Molecular differentiation of historic phage-type 80/81 and contemporary epidemic *Staphylococcus aureus*. *Proc Natl Acad Sci U S A,* 108**,** 18091-6.

DELEO, F. R., OTTO, M., KREISWIRTH, B. N. & CHAMBERS, H. F. 2010. Community-associated meticillin-resistant *Staphylococcus aureus*. *Lancet,* 375**,** 1557-68.

DEVRIES, A. S., LESHER, L., SCHLIEVERT, P. M., ROGERS, T., VILLAUME, L. G., DANILA, R. & LYNFIELD, R. 2011. Staphylococcal toxic shock syndrome 2000-2006: epidemiology, clinical features, and molecular characteristics. *PLoS One,* 6**,** e22997.

DIEP, B. A., CARLETON, H. A., CHANG, R. F., SENSABAUGH, G. F. & PERDREAU-REMINGTON, F. 2006. Roles of 34 virulence genes in the evolution of hospital- and community-associated strains of methicillin-resistant *Staphylococcus aureus*. *J Infect Dis,* 193**,** 1495-503.

DIEP, B. A., CHAMBERS, H. F., GRABER, C. J., SZUMOWSKI, J. D., MILLER, G., HAN, L. L., CHEN, J. H., LIN, F., LIN, J., PHAN, T. H., CARLETON, H. A., MCDOUGAL, L. K., TENOVER, F. C., COHEN, D. E., MAYER, K. H., SENSABAUGH, G. F. & PERDREAU-REMINGTON, F. 2008. Emergence of multidrug-resistant, community-associated, methicillin-resistant *Staphylococcus aureus* clone USA300 in men who have sex with men. *Ann Intern Med,* 148**,** 249-57.

DIEP, B. A., CHAN, L., TATTEVIN, P., KAJIKAWA, O., MARTIN, T. R., BASUINO, L., MAI, T. T., MARBACH, H., BRAUGHTON, K. R., WHITNEY, A. R., GARDNER, D. J., FAN, X., TSENG, C. W., LIU, G. Y., BADIOU, C., ETIENNE, J., LINA, G., MATTHAY, M. A., DELEO, F. R. & CHAMBERS, H. F. 2010. Polymorphonuclear leukocytes mediate *Staphylococcus aureus* Panton-Valentine leukocidin-induced lung inflammation and injury. *Proc Natl Acad Sci U S A,* 107**,** 5587-92.

DIEP, B. A., SENSABAUGH, G. F., SOMBOONNA, N., SOMBOONA, N. S., CARLETON, H. A. & PERDREAU-REMINGTON, F. 2004. Widespread skin and soft-tissue infections due to two methicillin-resistant *Staphylococcus aureus* strains harboring the genes for Panton-Valentine leucocidin. *J Clin Microbiol,* 42**,** 2080-4.

DOLEANS, A., AURELL, H., REYROLLE, M., LINA, G., FRENEY, J., VANDENESCH, F., ETIENNE, J. & JARRAUD, S. 2004. Clinical and environmental distributions of Legionella strains in France are different. *J Clin Microbiol,* 42**,** 458-60.

DONAHUE, J. A. & BALDWIN, J. N. 1966. Hemolysin and leukocidin production by 80/81 strains of *Staphylococcus aureus*. *J Infect Dis,* 116**,** 324-8.

DONKER, T., WALLINGA, J. & GRUNDMANN, H. 2010. Patient referral patterns and the spread of hospital-acquired infections through national health care

networks. *PLoS Comput Biol,* 6**,** e1000715.

DONKER, T., WALLINGA, J., SLACK, R. & GRUNDMANN, H. 2012. Hospital networks and the dispersal of hospital-acquired pathogens by patient transfer. *PLoS One,* 7**,** e35002.

DRUMMOND, A. J., HO, S. Y., PHILLIPS, M. J. & RAMBAUT, A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol,* 4**,** e88.

DRUMMOND, A. J., PYBUS, O. G., RAMBAUT, A., FORSBERG, R. & RODRIGO, A. G. 2003. Measurably evolving populations. *Trends Ecol Evol,* 18**,** 481-488.

DRUMMOND, A. J. & RAMBAUT, A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol,* 7**,** 214.

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., 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**,** 1008-15.

ENRIGHT, M. C. & SPRATT, B. G. 1998. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. *Microbiology,* 144 (Pt 11)**,** 3049-60.

ENTENZA, J. M., MOREILLON, P., SENN, M. M., KORMANEC, J., DUNMAN,

P. M., BERGER-BACHI, B., PROJAN, S. & BISCHOFF, M. 2005. Role of sigmaB in the expression of *Staphylococcus aureus* cell wall adhesins ClfA and FnbA and contribution to infectivity in a rat model of experimental endocarditis. *Infect Immun,* 73**,** 990-8.

ESPADINHA, D., FARIA, N. A., MIRAGAIA, M., LITO, L. M., MELO-CRISTINO, J., DE LENCASTRE, H. & MEDICOS SENTINELA, N. 2013. Extensive dissemination of methicillin-resistant *Staphylococcus aureus* (MRSA) between the hospital and the community in a country with a high prevalence of nosocomial MRSA. *PLoS One,* 8**,** e59960.

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, e001124.

FEIL, E. J., COOPER, J. E., GRUNDMANN, H., ROBINSON, D. A., ENRIGHT, M. C., BERENDT, T., PEACOCK, S. J., SMITH, J. M., MURPHY, M., SPRATT, B. G., MOORE, C. E. & DAY, N. P. J. 2003. How Clonal Is *Staphylococcus aureus*? *J Bacteriol,* 185**,** 3307-3316.

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**,** 1520-2.

FIELDS, B. S., BENSON, R. F. & BESSER, R. E. 2002. *Legionella* and Legionnaires' disease: 25 years of investigation. *Clin Microbiol Rev,* 15**,** 506-

26.

FISCHETTI, V. A., NOVICK, R. P., FERRETTI, J. J., PORTNOY, D. A. & ROOD, J. I. 2006. Gram-Positive Pathogens, Amer Society for Microbiology.

FISMAN, D. N., LIM, S., WELLENIUS, G. A., JOHNSON, C., BRITZ, P., GASKINS, M., MAHER, J., MITTLEMAN, M. A., SPAIN, C. V., HAAS, C. N. & NEWBERN, C. 2005. It's not the heat, it's the humidity: wet weather increases legionellosis risk in the greater Philadelphia metropolitan area. *J Infect Dis,* 192**,** 2066-73.

FITZGERALD, J. R., STURDEVANT, D. E., MACKIE, S. M., GILL, S. R. & MUSSER, J. M. 2001. Evolutionary genomics of *Staphylococcus aureus*: insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. *Proc Natl Acad Sci U S A,* 98**,** 8821-6.

FLIERMANS, C. B., CHERRY, W. B., ORRISON, L. H. & THACKER, L. 1979. Isolation of *Legionella pneumophila* from nonepidemic-related aquatic habitats. *Appl Environ Microbiol,* 37**,** 1239-42.

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**,** 482-6.

FRANK, C., WERBER, D., CRAMER, J. P., ASKAR, M., FABER, M., AN DER HEIDEN, M., BERNARD, H., FRUTH, A., PRAGER, R., SPODE, A., WADL, M., ZOUFALY, A., JORDAN, S., KEMPER, M. J., FOLLIN, P., MULLER, L., KING, L. A., ROSNER, B., BUCHHOLZ, U., STARK, K., KRAUSE, G. & TEAM, H. U. S. I. 2011. Epidemic profile of Shiga-toxin-producing *Escherichia coli* O104:H4 outbreak in Germany. *N Engl J Med,*

365**,** 1771-80.

FRANK, D. N., FEAZEL, L. M., BESSESEN, M. T., PRICE, C. S., JANOFF, E. N. & PACE, N. R. 2010. The human nasal microbiota and *Staphylococcus aureus* carriage. *PLoS One,* 5**,** e10598.

FRASER, D. W., DEUBNER, D. C., HILL, D. L. & GILLIAM, D. K. 1979. Nonpneumonic, short-incubation-period Legionellosis (Pontiac fever) in men who cleaned a steam turbine condenser. *Science,* 205**,** 690-1.

FRASER, D. W., TSAI, T. R., ORENSTEIN, W., PARKIN, W. E., BEECHAM, H. J., SHARRAR, R. G., HARRIS, J., MALLISON, G. F., MARTIN, S. M., MCDADE, J. E., SHEPARD, C. C. & BRACHMAN, P. S. 1977. Legionnaires' disease: description of an epidemic of pneumonia. *N Engl J Med,* 297**,** 1189-97.

FRONZES, R., CHRISTIE, P. J. & WAKSMAN, G. 2009. The structural biology of type IV secretion systems. *Nat Rev Microbiol,* 7**,** 703-14.

GAIA, V., FRY, N. K., AFSHAR, B., LUCK, P. C., MEUGNIER, H., ETIENNE, J., PEDUZZI, R. & HARRISON, T. G. 2005. Consensus sequence-based scheme for epidemiological typing of clinical and environmental isolates of *Legionella pneumophila*. *J Clin Microbiol,* 43**,** 2047-52.

GAO, L. Y. & ABU KWAIK, Y. 1999. Apoptosis in macrophages and alveolar epithelial cells during early stages of infection by *Legionella pneumophila* and its role in cytopathogenicity. *Infect Immun,* 67**,** 862-70.

GAO, W., CHUA, K., DAVIES, J. K., NEWTON, H. J., SEEMANN, T., HARRISON, P. F., HOLMES, N. E., RHEE, H. W., HONG, J. I., HARTLAND, E. L., STINEAR, T. P. & HOWDEN, B. P. 2010. Two novel point mutations in clinical *Staphylococcus aureus* reduce linezolid

susceptibility and switch on the stringent response to promote persistent infection. *PLoS Pathog,* 6**,** e1000944.

GARCIA-ALVAREZ, L., HOLDEN, M. T., LINDSAY, H., WEBB, C. R., BROWN, D. F., CURRAN, M. D., WALPOLE, E., BROOKS, K., PICKARD, D. J., TEALE, C., PARKHILL, J., BENTLEY, S. D., EDWARDS, G. F., GIRVAN, E. K., KEARNS, A. M., PICHON, B., HILL, R. L., LARSEN, A. R., SKOV, R. L., PEACOCK, S. J., MASKELL, D. J. & HOLMES, M. A. 2011. Meticillin-resistant *Staphylococcus aureus* with a novel *mecA* homologue in human and bovine populations in the UK and Denmark: a descriptive study. *Lancet Infect Dis,* 11**,** 595-603.

GARDETE, S., WU, S. W., GILL, S. & TOMASZ, A. 2006. Role of VraSR in antibiotic resistance and antibiotic-induced stress response in *Staphylococcus aureus*. *Antimicrob Agents Chemother,* 50**,** 3424-34.

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. *N Engl J Med,* 364**,** 730-9.

GILLET, Y., ETIENNE, J., LINA, G. & VANDENESCH, F. 2008. Association of necrotizing pneumonia with Panton-Valentine leukocidin-producing *Staphylococcus aureus*, regardless of methicillin resistance. *Clin Infect Dis,* 47**,** 985-6.

GLOCKNER, G., ALBERT-WEISSENBERGER, C., WEINMANN, E., JACOBI, S., SCHUNDER, E., STEINERT, M., HACKER, J. & HEUNER, K. 2008. Identification and characterization of a new conjugation/type IVA secretion system (*trb*/*tra*) of *Legionella pneumophila* Corby localized on two mobile

genomic islands. *Int J Med Microbiol,* 298**,** 411-28.

GOERKE, C., MATIAS Y PAPENBERG, S., DASBACH, S., DIETZ, K., ZIEBACH, R., KAHL, B. C. & WOLZ, C. 2004. Increased frequency of genomic alterations in *Staphylococcus aureus* during chronic infection is in part due to phage mobilization. *J Infect Dis,* 189**,** 724-34.

GOERKE, C. & WOLZ, C. 2010. Adaptation of *Staphylococcus aureus* to the cystic fibrosis lung. *Int J Med Microbiol,* 300**,** 520-5.

GOLDIE, D. J., ALDER, V. G. & GILLESPIE, W. A. 1971. Changes in the drug resistance of *Staphylococcus aureus* in a non-hospital population during a 20-year period. *J Clin Pathol,* 24**,** 44-7.

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., DIDELOT, X., PETO, T. E., 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.

GOLUBCHIK, T., BRUEGGEMANN, A. B., STREET, T., GERTZ, R. E., JR., SPENCER, C. C., HO, T., GIANNOULATOU, E., LINK-GELLES, R., HARDING, R. M., BEALL, B., PETO, T. E., MOORE, M. R., DONNELLY, P., CROOK, D. W. & BOWDEN, R. 2012. Pneumococcal genome sequencing tracks a vaccine escape variant formed through a multi-fragment recombination event. *Nat Genet,* 44**,** 352-5.

GOMEZ-VALERO, L., RUSNIOK, C., CAZALET, C. & BUCHRIESER, C. 2011. Comparative and functional genomics of *Legionella* identified eukaryotic like proteins as key players in host-pathogen interactions. *Front Microbiol,* 2**,** 208.

GORDON, A. 2010. FASTX-toolkit. http://hannonlab.cshl.edu/fastx_toolkit/

GOUET, P., COURCELLE, E., STUART, D. I. & METOZ, F. 1999. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics,* 15**,** 305-8.

GOULD, J. C. 1958. Environmental Penicillin and Penicillin-Resistant *Staphylococcus aureus*. *Lancet,* 271**,** 489-493.

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. *Mol Biol Evol,* 28**,** 1593-603.

GRUNDMANN, H., AIRES-DE-SOUSA, M., BOYCE, J. & TIEMERSMA, E. 2006. Emergence and resurgence of meticillin-resistant *Staphylococcus aureus* as a public-health threat. *Lancet,* 368**,** 874-885.

GUBLER, D. J. 2011. Dengue, Urbanization and Globalization: The Unholy Trinity of the 21(st) Century. *Trop Med Health,* 39**,** 3-11.

GUINANE, C. M., BEN ZAKOUR, N. L., TORMO-MAS, M. A., WEINERT, L. A., LOWDER, B. V., CARTWRIGHT, R. A., SMYTH, D. S., SMYTH, C. J., LINDSAY, J. A., GOULD, K. A., WITNEY, A., HINDS, J., BOLLBACK, J. P., RAMBAUT, A., PENADES, J. R. & FITZGERALD, J. R. 2010. Evolutionary genomics of *Staphylococcus aureus* reveals insights into the origin and molecular basis of ruminant host adaptation. *Genome Biol Evol,* 2**,** 454-66.

GYSIN, C., ALOTHMAN, G. A. & PAPSIN, B. C. 2000. Sinonasal disease in cystic

fibrosis: clinical characteristics, diagnosis, and management. *Pediatr Pulmonol,* 30**,** 481-9.

HACKBARTH, C. J. & CHAMBERS, H. F. 1993. *blaI* and *blaR1* regulate beta-lactamase and PBP2a production in methicillin-resistant *Staphylococcus aureus*. *Antimicrob Agents Chemother,* 37**,** 1144-9.

HANIQUE, S., COLOMBO, M. L., GOORMAGHTIGH, E., SOUMILLION, P., FRERE, J. M. & JORIS, B. 2004. Evidence of an intramolecular interaction between the two domains of the BlaR1 penicillin receptor during the signal transduction. *J Biol Chem,* 279**,** 14264-72.

HANSSEN, A. M. & ERICSON SOLLID, J. U. 2006. SCCmec in staphylococci: genes on the move. *FEMS Immunol Med Microbiol,* 46**,** 8-20.

HARDING, C. R., SCHROEDER, G. N., REYNOLDS, S., KOSTA, A., COLLINS, J. W., MOUSNIER, A. & FRANKEL, G. 2012. *Legionella pneumophila* pathogenesis in the *Galleria mellonella* infection model. *Infect Immun,* 80**,** 2780-90.

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**,** 130-6.

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 LENCASTRE, H., PARKHILL, J., PEACOCK, S. J. & BENTLEY, S. D. 2010. Evolution of MRSA during hospital transmission and intercontinental spread. *Science,* 327**,** 469-74.

HARRISON, F. 2007. Microbial ecology of the cystic fibrosis lung. *Microbiology*, 153**,** 917-23.

HEALTH PROTECTION AGENCY. 2012. Healthcare-Associated Infection and Antimicrobial Resistance: 2010-2011. Online report.

HEALTH PROTECTION AGENCY. 2012. Results from the mandatory surveillance of MRSA bacteraemia. Online report.

HEINSEN, A., BENDTSEN, F. & FOMSGAARD, A. 2000. A phylogenetic analysis elucidating a case of patient-to-patient transmission of hepatitis C virus during surgery. *J Hosp Infect,* 46**,** 309-13.

HENDRIKSEN, R. S., PRICE, L. B., SCHUPP, J. M., GILLECE, J. D., KAAS, R. S., ENGELTHALER, D. M., BORTOLAIA, V., PEARSON, T., WATERS, A. E., UPADHYAY, B. P., SHRESTHA, S. D., ADHIKARI, S., SHAKYA, G., KEIM, P. S. & AARESTRUP, F. M. 2011. Population genetics of *Vibrio cholerae* from Nepal in 2010: evidence on the origin of the Haitian outbreak. *MBio,* 2**,** e00157-11.

HEROLD, B. C., IMMERGLUCK, L. C., MARANAN, M. C., LAUDERDALE, D. S., GASKIN, R. E., BOYLE-VAVRA, S., LEITCH, C. D. & DAUM, R. S. 1998. Community-acquired methicillin-resistant *Staphylococcus aureus* in children with no identified predisposing risk. *JAMA,* 279**,** 593-8.

HEUNER, K. & SWANSON, M. 2008. *Legionella*: Molecular Microbiology, Caister Academic Press.

HIRAMATSU, K., ARITAKA, N., HANAKI, H., KAWASAKI, S., HOSODA, Y., HORI, S., FUKUCHI, Y. & KOBAYASHI, I. 1997. Dissemination in Japanese hospitals of strains of *Staphylococcus aureus* heterogeneously resistant to vancomycin. *Lancet,* 350**,** 1670-1673.

HIRON, A., FALORD, M., VALLE, J., DEBARBOUILLE, M. & MSADEK, T. 2011. Bacitracin and nisin resistance in *Staphylococcus aureus*: a novel pathway involving the BraS/BraR two-component system (SA2417/SA2418) and both the BraD/BraE and VraD/VraE ABC transporters. *Mol Microbiol,* 81**,** 602-22.

HO, P. L., CHOW, K. H., LO, P. Y., LEE, K. F. & LAI, E. L. 2009. Changes in the epidemiology of methicillin-resistant *Staphylococcus aureus* associated with spread of the ST45 lineage in Hong Kong. *Diagn Microbiol Infect Dis,* 64**,** 131-7.

HOGARDT, M. & HEESEMANN, J. 2010. Adaptation of *Pseudomonas aeruginosa* during persistence in the cystic fibrosis lung. *Int J Med Microbiol,* 300**,** 557-62.

HOLDEN, M. T., FEIL, E. J., LINDSAY, J. A., PEACOCK, S. J., DAY, N. P., ENRIGHT, M. C., FOSTER, T. J., MOORE, C. E., HURST, L., ATKIN, R., BARRON, A., BASON, N., BENTLEY, S. D., CHILLINGWORTH, C., CHILLINGWORTH, T., CHURCHER, C., CLARK, L., CORTON, C., CRONIN, A., DOGGETT, J., DOWD, L., FELTWELL, T., HANCE, Z., HARRIS, B., HAUSER, H., HOLROYD, S., JAGELS, K., JAMES, K. D., LENNARD, N., LINE, A., MAYES, R., MOULE, S., MUNGALL, K., ORMOND, D., QUAIL, M. A., RABBINOWITSCH, E., RUTHERFORD, K., SANDERS, M., SHARP, S., SIMMONDS, M., STEVENS, K., WHITEHEAD, S., BARRELL, B. G., SPRATT, B. G. & PARKHILL, J. 2004. Complete genomes of two clinical *Staphylococcus aureus* strains: evidence for the rapid evolution of virulence and drug resistance. *Proc Natl Acad Sci U S A,* 101**,** 9786-91.

HOLDEN, M. T., HSU, L. Y., KURT, K., WEINERT, L. A., MATHER, A. E., HARRIS, S. R., STROMMENGER, B., LAYER, F., WITTE, W., DE

LENCASTRE, 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**,** 653-64.

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**,** 888-92.

HOLT, D. C., HOLDEN, M. T., TONG, S. Y., 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 Biol Evol,* 3**,** 881-95.

HUGOSSON, A., HJORTH, M., BERNANDER, S., CLAESSON, B. E., JOHANSSON, A., LARSSON, H., NOLSKOG, P., PAP, J., SVENSSON, N. & ULLERYD, P. 2007. A community outbreak of Legionnaires' disease from an industrial cooling tower: assessment of clinical features and diagnostic procedures. *Scand J Infect Dis,* 39**,** 217-24.

HUSON, D. H. & BRYANT, D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol,* 23**,** 254-67.

JANSEN VAN RENSBURG, M. J., ELIYA MADIKANE, V., WHITELAW, A., CHACHAGE, M., HAFFEJEE, S. & GAY ELISHA, B. 2011. The dominant methicillin-resistant *Staphylococcus aureus* clone from hospitals in Cape Town has an unusual genotype: ST612. *Clin Microbiol Infect,* 17**,** 785-92.

JEVONS, M. P. 1961. "Celbenin"-resistant Staphylococci. *BMJ,* 1**,** 113-4.

JOHNSON, A. P., PEARSON, A. & DUCKWORTH, G. 2005. Surveillance and epidemiology of MRSA bacteraemia in the UK. *J Antimicrob Chemother,* 56**,** 455-62.

JONES, R. N. 2003. Global epidemiology of antimicrobial resistance among community-acquired and nosocomial pathogens: a five-year summary from the SENTRY Antimicrobial Surveillance Program (1997-2001). *Semin Respir Crit Care Med,* 24**,** 121-34.

JUHAS, M., CROOK, D. W., DIMOPOULOU, I. D., LUNTER, G., HARDING, R. M., FERGUSON, D. J. & HOOD, D. W. 2007. Novel type IV secretion system involved in propagation of genomic islands. *J Bacteriol,* 189**,** 761-71.

KAHL, B. C. 2010. Impact of *Staphylococcus aureus* on the pathogenesis of chronic cystic fibrosis lung disease. *Int J Med Microbiol,* 300**,** 514-9.

KAITO, C., SAITO, Y., IKUO, M., OMAE, Y., MAO, H., NAGANO, G., FUJIYUKI, T., NUMATA, S., HAN, X., OBATA, K., HASEGAWA, S., YAMAGUCHI, H., INOKUCHI, K., ITO, T., HIRAMATSU, K. & SEKIMIZU, K. 2013. Mobile genetic element SCC*mec*-encoded psm-*mec* RNA suppresses translation of *agrA* and attenuates MRSA virulence. *PLoS Pathog,* 9**,** e1003269.

KARLSSON-KANTH, A., TEGMARK-WISELL, K., ARVIDSON, S. & OSCARSSON, J. 2006. Natural human isolates of *Staphylococcus aureus*

selected for high production of proteases and alpha-hemolysin are sigmaB deficient. *Int J Med Microbiol,* 296**,** 229-36.

KATZ, S. M., HABIB, W. A., HAMMEL, J. M. & NASH, P. 1982. Lack of airborne spread of infection by *Legionella pneumophila* among guinea pigs. *Infect Immun,* 38**,** 620-2.

KELLEY, D. R., SCHATZ, M. C. & SALZBERG, S. L. 2010. Quake: quality-aware detection and correction of sequencing errors. *Genome Biol,* 11**,** R116.

KEREM, B., ROMMENS, J. M., BUCHANAN, J. A., MARKIEWICZ, D., COX, T. K., CHAKRAVARTI, A., BUCHWALD, M. & TSUI, L. C. 1989. Identification of the cystic fibrosis gene: genetic analysis. *Science,* 245**,** 1073-80.

KLEIN, E., SMITH, D. L. & LAXMINARAYAN, R. 2007. Hospitalizations and deaths caused by methicillin-resistant *Staphylococcus aureus*, United States, 1999-2005. *Emerg Infect Dis,* 13**,** 1840-6.

KLUYTMANS, J., VAN BELKUM, A. & VERBRUGH, H. 1997. Nasal carriage of *Staphylococcus aureus*: epidemiology, underlying mechanisms, and associated risks. *Clin Microbiol Rev,* 10**,** 505-20.

KNAPP, J. S. & HOOK, E. W., 3RD 1988. Prevalence and persistence of *Neisseria cinerea* and other *Neisseria* spp. in adults. *J Clin Microbiol,* 26**,** 896-900.

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**,** 2514-22.

KOCK, R., BECKER, K., COOKSON, B., VAN GEMERT-PIJNEN, J. E.,

HARBARTH, S., KLUYTMANS, J., MIELKE, M., PETERS, G., SKOV, R. L., STRUELENS, M. J., TACCONELLI, E., NAVARRO TORNE, A., WITTE, W. & FRIEDRICH, A. W. 2010. Methicillin-resistant *Staphylococcus aureus* (MRSA): burden of disease and control challenges in Europe. *Euro Surveill,* 15**,** 19688.

KOREEN, L., RAMASWAMY, S. V., GRAVISS, E. A., NAIDICH, S., MUSSER, J. M. & KREISWIRTH, B. N. 2004. *spa* typing method for discriminating among *Staphylococcus aureus* isolates: implications for use of a single marker to detect genetic micro- and macrovariation. *J Clin Microbiol,* 42**,** 792-9.

KOREN, S., HARHAY, G. P., SMITH, T. P., BONO, J. L., HARHAY, D. M., MCVEY, S. D., RADUNE, D., BERGMAN, N. H. & PHILLIPPY, A. M. 2013. Reducing assembly complexity of microbial genomes with single-molecule sequencing. *Genome Biol,* 14**,** R101.

KOS, V. N., DESJARDINS, C. A., GRIGGS, A., CERQUEIRA, G., VAN TONDER, A., HOLDEN, M. T., GODFREY, P., PALMER, K. L., BODI, K., MONGODIN, E. F., WORTMAN, J., FELDGARDEN, M., LAWLEY, T., GILL, S. R., HAAS, B. J., BIRREN, B. & GILMORE, M. S. 2012. Comparative genomics of vancomycin-resistant *Staphylococcus aureus* strains and their positions within the clade most commonly associated with Methicillin-resistant *S. aureus* hospital-acquired infection in the United States. *MBio,* 3.

KOSAKOVSKY POND, S. L., FROST, S. D. & MUSE, S. V. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics,* 21**,** 676-9.

KOSAKOVSKY POND, S. L., POSADA, D., GRAVENOR, M. B., WOELK, C. H. & FROST, S. D. 2006. GARD: a genetic algorithm for recombination detection. *Bioinformatics,* 22**,** 3096-8.

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. 2012. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *N Engl J Med,* 366**,** 2267-75.

KUMARASAMY, K. K., TOLEMAN, M. A., WALSH, T. R., BAGARIA, J., BUTT, F., BALAKRISHNAN, R., CHAUDHARY, U., DOUMITH, M., GISKE, C. G., IRFAN, S., KRISHNAN, P., KUMAR, A. V., MAHARJAN, S., MUSHTAQ, S., NOORIE, T., PATERSON, D. L., PEARSON, A., PERRY, C., PIKE, R., RAO, B., RAY, U., SARMA, J. B., SHARMA, M., SHERIDAN, E., THIRUNARAYAN, M. A., TURTON, J., UPADHYAY, S., WARNER, M., WELFARE, W., LIVERMORE, D. M. & WOODFORD, N. 2010. Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study. *Lancet Infect Dis,* 10**,** 597-602.

KWAN, T., LIU, J., DUBOW, M., GROS, P. & PELLETIER, J. 2005. The complete genomes and proteomes of 27 *Staphylococcus aureus* bacteriophages. *Proc Natl Acad Sci U S A,* 102**,** 5174-9.

LARKIN, M. A., BLACKSHIELDS, G., BROWN, N. P., CHENNA, R., MCGETTIGAN, P. A., MCWILLIAM, H., VALENTIN, F., WALLACE, I. M., WILM, A., LOPEZ, R., THOMPSON, J. D., GIBSON, T. J. & HIGGINS, D. G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics,* 23**,** 2947-8.

LARSON, H. E., PRICE, A. B., HONOUR, P. & BORRIELLO, S. P. 1978. *Clostridium difficile* and the aetiology of pseudomembranous colitis. *Lancet,*

1**,** 1063-6.

LAUTNER, M., SCHUNDER, E., HERRMANN, V. & HEUNER, K. 2013. Regulation, integrase-dependent excision, and horizontal transfer of genomic islands in *Legionella pneumophila*. *J Bacteriol,* 195**,** 1583-97.

LEMEY, P., KOSAKOVSKY POND, S. L., DRUMMOND, A. J., PYBUS, O. G., SHAPIRO, B., BARROSO, H., TAVEIRA, N. & RAMBAUT, A. 2007. Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics. *PLoS Comput Biol,* 3**,** e29.

LEMEY, P., RAMBAUT, A., BEDFORD, T., FARIA, N. R., BIELEJEC, F., BAELE, G., RUSSELL, C. A., SMITH, D. J., PYBUS, O. G., BROCKMANN, D. & SUCHARD, M. A. 2012. The seasonal flight of influenza: a unified framework for spatiotemporal hypothesis testing. 16-16.

LEMEY, P., RAMBAUT, A., DRUMMOND, A. J. & SUCHARD, M. A. 2009. Bayesian phylogeography finds its roots. *PLoS Comput Biol,* 5**,** e1000520.

LESKI, T. A. & TOMASZ, A. 2005. Role of penicillin-binding protein 2 (PBP2) in the antibiotic susceptibility and cell wall cross-linking of *Staphylococcus aureus*: evidence for the cooperative functioning of PBP2, PBP4, and PBP2A. *J Bacteriol,* 187**,** 1815-24.

LEVDIKOV, V. M., BLAGOVA, E., JOSEPH, P., SONENSHEIN, A. L. & WILKINSON, A. J. 2006. The structure of CodY, a GTP- and isoleucine-responsive regulator of stationary phase and virulence in gram-positive bacteria. *J Biol Chem,* 281**,** 11366-73.

LEWIS, T., LOMAN, N. J., BINGLE, L., JUMAA, P., WEINSTOCK, G. M., MORTIBOY, D. & PALLEN, M. J. 2010. High-throughput whole-genome sequencing to dissect the epidemiology of Acinetobacter baumannii isolates

from a hospital outbreak. *J Hosp Infect,* 75**,** 37-41.

LI, H. & DURBIN, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* 25**,** 1754-60.

LI, H. & DURBIN, R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics,* 26**,** 589-95.

LI, H., RUAN, J. & DURBIN, R. 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res,* 18**,** 1851-8.

LI, J. S., O'BRIEN, E. D. & GUEST, C. 2002. A review of national legionellosis surveillance in Australia, 1991 to 2000. *Commun Dis Intell Q Rep,* 26**,** 461-8.

LIEBERMAN, T. D., MICHEL, J. B., AINGARAN, M., POTTER-BYNOE, G., ROUX, D., DAVIS, M. R., JR., SKURNIK, D., LEIBY, N., LIPUMA, J. J., GOLDBERG, J. B., MCADAM, A. J., PRIEBE, G. P. & KISHONY, R. 2011. Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nat Genet,* 43**,** 1275-80.

LINDSAY, J. A. 2008. *Staphylococcus:* molecular genetics, Norfolk, Caister Academic Press.

LINDSAY, J. A. 2013. Hospital-associated MRSA and antibiotic resistance-what have we learned from genomics? *Int J Med Microbiol,* 303**,** 318-23.

LINDSAY, J. A. & HOLDEN, M. T. 2006. Understanding the rise of the superbug: investigation of the evolution and genomic variation of *Staphylococcus aureus*. *Funct Integr Genomics,* 6**,** 186-201.

LIU, G. Y., ESSEX, A., BUCHANAN, J. T., DATTA, V., HOFFMAN, H. M.,

BASTIAN, J. F., FIERER, J. & NIZET, V. 2005. *Staphylococcus aureus* golden pigment impairs neutrophil killing and promotes virulence through its antioxidant activity. *J Exp Med,* 202**,** 209-15.

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.

LOMAN, N. J., CONSTANTINIDOU, C., CHRISTNER, M., ROHDE, H., CHAN, J. Z., QUICK, J., WEIR, J. C., QUINCE, C., SMITH, G. P., BETLEY, J. R., AEPFELBACHER, M. & PALLEN, M. J. 2013. A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of Shiga-toxigenic *Escherichia coli* O104:H4. *JAMA,* 309**,** 1502-10.

LOMAN, N. J., MISRA, R. V., DALLMAN, T. J., CONSTANTINIDOU, C., GHARBIA, S. E., WAIN, J. & PALLEN, M. J. 2012. Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol,* 30**,** 434-9.

LOMMA, M., GOMEZ VALERO, L., RUSNIOK, C. & BUCHRIESER, C. 2009. *Legionella pneumophila* - Host Interactions: Insights Gained from Comparative Genomics and Cell Biology. *Genome Dyn,* 6**,** 170-86.

LOWDER, B. V., GUINANE, C. M., BEN ZAKOUR, N. L., WEINERT, L. A., CONWAY-MORRIS, A., CARTWRIGHT, R. A., SIMPSON, A. J., RAMBAUT, A., NUBEL, U. & FITZGERALD, J. R. 2009. Recent human-to-poultry host jump, adaptation, and pandemic spread of *Staphylococcus aureus*. *Proc Natl Acad Sci U S A,* 106**,** 19545-50.

LOWY, F. D. 1998. *Staphylococcus aureus* infections. *N Engl J Med,* 339**,** 520-32.

LUCK, C., FRY, N. K., HELBIG, J. H., JARRAUD, S. & HARRISON, T. G. 2013. Typing methods for *Legionella*. *Methods Mol Biol,* 954**,** 119-48.

LUO, Z. Q. & ISBERG, R. R. 2004. Multiple substrates of the *Legionella pneumophila* Dot/Icm system identified by interbacterial protein transfer. *Proc Natl Acad Sci U S A,* 101**,** 841-6.

MA, X. X., ITO, T., KONDO, Y., CHO, M., YOSHIZAWA, Y., KANEKO, J., KATAI, A., HIGASHIIDE, M., LI, S. & HIRAMATSU, K. 2008. Two different Panton-Valentine leukocidin phage lineages predominate in Japan. *J Clin Microbiol,* 46**,** 3246-58.

MA, X. X., ITO, T., TIENSASITORN, C., JAMKLANG, M., CHONGTRAKOOL, P., BOYLE-VAVRA, S., DAUM, R. S. & HIRAMATSU, K. 2002. Novel type of staphylococcal cassette chromosome *mec* identified in community-acquired methicillin-resistant *Staphylococcus aureus* strains. *Antimicrob Agents Chemother,* 46**,** 1147-52.

MAIDEN, M. C., BYGRAVES, J. A., FEIL, E., MORELLI, G., RUSSELL, J. E., URWIN, R., ZHANG, Q., ZHOU, J., ZURTH, K., CAUGANT, D. A., FEAVERS, I. M., ACHTMAN, M. & SPRATT, B. G. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA,* 95**,** 3140-5.

MAINZ, J. G., NAEHRLICH, L., SCHIEN, M., KADING, M., SCHILLER, I., MAYR, S., SCHNEIDER, G., WIEDEMANN, B., WIEHLMANN, L., CRAMER, N., PFISTER, W., KAHL, B. C., BECK, J. F. & TUMMLER, B. 2009. Concordant genotype of upper and lower airways *P. aeruginosa* and *S. aureus* isolates in cystic fibrosis. *Thorax,* 64**,** 535-40.

MAJERCZYK, C. D., DUNMAN, P. M., LUONG, T. T., LEE, C. Y., SADYKOV,

M. R., SOMERVILLE, G. A., BODI, K. & SONENSHEIN, A. L. 2010. Direct targets of CodY in *Staphylococcus aureus*. *J Bacteriol,* 192**,** 2861-77.

MAREE, C. L., DAUM, R. S., BOYLE-VAVRA, S., MATAYOSHI, K. & MILLER, L. G. 2007. Community-associated methicillin-resistant *Staphylococcus aureus* isolates causing healthcare-associated infections. *Emerg Infect Dis,* 13**,** 236-42.

MARPLES, R. R. & REITH, S. 1992. Methicillin-resistant *Staphylococcus aureus* in England and Wales. *Commun Dis Rep CDR Wkly.,* 2**,** R25-9.

MARSHALL, B. J. & WARREN, J. R. 1984. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet,* 1**,** 1311-5.

MARSTON, B. J., LIPMAN, H. B. & BREIMAN, R. F. 1994. Surveillance for Legionnaires' disease. Risk factors for morbidity and mortality. *Arch Intern Med,* 154**,** 2417-22.

MARTIN, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal,* 17**,** 10-12.

MARTTINEN, P., HANAGE, W. P., CROUCHER, N. J., CONNOR, T. R., HARRIS, S. R., BENTLEY, S. D. & CORANDER, J. 2012. Detection of recombination events in bacterial genomes from large population samples. *Nucleic Acids Res,* 40**,** e6.

MASTRO, T. D., FIELDS, B. S., BREIMAN, R. F., CAMPBELL, J., PLIKAYTIS, B. D. & SPIKA, J. S. 1991. Nosocomial Legionnaires' disease and use of medication nebulizers. *J Infect Dis,* 163**,** 667-71.

MCADAM, P. R., HOLMES, A., TEMPLETON, K. E. & FITZGERALD, J. R. 2011. Adaptive evolution of *Staphylococcus aureus* during chronic

endobronchial infection of a cystic fibrosis patient. *PLoS One,* 6**,** e24301.

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**,** 9107-12.

MCCARTHY, A. J. & LINDSAY, J. A. 2010. Genetic variation in *Staphylococcus aureus* surface and immune evasion genes is lineage associated: implications for vaccine design and host-pathogen interactions. *BMC Microbiol,* 10**,** 173.

MCCORMICK, D., THORN, S., MILNE, D., EVANS, C., STEVENSON, J., LLANO, M., DONAGHY, M. & TEAM, I. M. 2012. Public health response to an outbreak of Legionnaires' disease in Edinburgh, United Kingdom, June 2012. *Euro Surveill,* 17.

MCDADE, J. E., SHEPARD, C. C., FRASER, D. W., TSAI, T. R., REDUS, M. A. & DOWDLE, W. R. 1977. Legionnaires' disease: isolation of a bacterium and demonstration of its role in other respiratory disease. *N Engl J Med,* 297**,** 1197-203.

MEIER, S., GOERKE, C., WOLZ, C., SEIDL, K., HOMEROVA, D., SCHULTHESS, B., KORMANEC, J., BERGER-BACHI, B. & BISCHOFF, M. 2007. sigmaB and the sigmaB-dependent *arlRS* and *yabJ-spoVG* loci affect capsule formation in *Staphylococcus aureus*. *Infect Immun,* 75**,** 4562-71.

METZKER, M. L. 2010. Sequencing technologies - the next generation. *Nat Rev*

Genet, 11**,** 31-46.

MIKHAIL, M., WANG, B., LEMEY, P., BECKTHOLD, B., VANDAMME, A. M., GILL, M. J. & SAKSENA, N. K. 2005. Role of viral evolutionary rate in HIV-1 disease progression in a linked cohort. *Retrovirology,* 2**,** 41.

MOISAN, H., BROUILLETTE, E., JACOB, C. L., LANGLOIS-BEGIN, P., MICHAUD, S. & MALOUIN, F. 2006. Transcription of virulence factors in *Staphylococcus aureus* small-colony variants isolated from cystic fibrosis patients is influenced by SigB. *J Bacteriol,* 188**,** 64-76.

MORRIS, G. K., PATTON, C. M., FEELEY, J. C., JOHNSON, S. E., GORMAN, G., MARTIN, W. T., SKALIY, P., MALLISON, G. F., POLITI, B. D. & MACKEL, D. C. 1979. Isolation of the Legionnaires' disease bacterium from environmental samples. *Ann Intern Med,* 90**,** 664-6.

MOURVILLIER, B., TROUILLET, J. L., TIMSIT, J. F., BAUDOT, J., CHASTRE, J., REGNIER, B., GIBERT, C. & WOLFF, M. 2004. Infective endocarditis in the intensive care unit: clinical spectrum and prognostic factors in 228 consecutive patients. *Intensive Care Med,* 30**,** 2046-52.

MURCHAN, S., AUCKEN, H. M., O'NEILL G, L., GANNER, M. & COOKSON, B. D. 2004. Emergence, spread, and characterization of phage variants of epidemic methicillin-resistant *Staphylococcus aureus* 16 in England and Wales. *J Clin Microbiol,* 42**,** 5154-60.

MUTREJA, A., KIM, D. W., THOMSON, N. R., CONNOR, T. R., LEE, J. H., KARIUKI, S., CROUCHER, N. J., CHOI, S. Y., HARRIS, S. R., LEBENS, M., NIYOGI, S. K., KIM, E. J., RAMAMURTHY, T., CHUN, J., WOOD, J. L., CLEMENS, J. D., CZERKINSKY, C., NAIR, G. B., HOLMGREN, J., PARKHILL, J. & DOUGAN, G. 2011. Evidence for several waves of global transmission in the seventh cholera pandemic. *Nature,* 477**,** 462-5.

MWANGI, M. M., WU, S. W., ZHOU, Y., SIERADZKI, K., DE LENCASTRE, H., RICHARDSON, P., BRUCE, D., RUBIN, E., MYERS, E., SIGGIA, E. D. & TOMASZ, A. 2007. Tracking the in vivo evolution of multidrug resistance in *Staphylococcus aureus* by whole-genome sequencing. *Proc Natl Acad Sci U S A,* 104**,** 9451-6.

NAGAEV, I., BJORKMAN, J., ANDERSSON, D. I. & HUGHES, D. 2001. Biological cost and compensatory evolution in fusidic acid-resistant *Staphylococcus aureus*. *Mol Microbiol,* 40**,** 433-9.

NAGAI, H., KAGAN, J. C., ZHU, X., KAHN, R. A. & ROY, C. R. 2002. A bacterial guanine nucleotide exchange factor activates ARF on *Legionella* phagosomes. *Science,* 295**,** 679-82.

NAHMIAS, A., SAKURAI, N., BLUMBERG, R., GOE GE, A. & SULZER, C. 1961. The *Staphylococcus* 80/81 complex: epidemiological and laboratory observations. *J Infect Dis,* 109**,** 211-22.

NCBI 2010. SRA Handbook, National Center for Biotechnology Information (US).

NELSON, M. I., LEMEY, P., TAN, Y., VINCENT, A., LAM, T. T., DETMER, S., VIBOUD, C., SUCHARD, M. A., RAMBAUT, A., HOLMES, E. C. & GRAMER, M. 2011. Spatial dynamics of human-origin H1 influenza A virus in North American swine. *PLoS Pathog,* 7**,** e1002077.

NGUYEN, T. M., ILEF, D., JARRAUD, S., ROUIL, L., CAMPESE, C., CHE, D., HAEGHEBAERT, S., GANIAYRE, F., MARCEL, F., ETIENNE, J. & DESENCLOS, J. C. 2006. A community-wide outbreak of Legionnaires disease linked to industrial cooling towers--how far can contaminated aerosols spread? *J Infect Dis,* 193**,** 102-11.

NIENABER, J. J., SHARMA KUINKEL, B. K., CLARKE-PEARSON, M., LAMLERTTHON, S., PARK, L., RUDE, T. H., BARRIERE, S., WOODS, C. W., CHU, V. H., MARIN, M., BUKOVSKI, S., GARCIA, P., COREY, G. R., KORMAN, T., DOCO-LECOMPTE, T., MURDOCH, D. R., RELLER, L. B., FOWLER, V. G., JR. & INTERNATIONAL COLLABORATION ON ENDOCARDITIS-MICROBIOLOGY, I. 2011. Methicillin-susceptible *Staphylococcus aureus* endocarditis isolates are associated with clonal complex 30 genotype and a distinct repertoire of enterotoxins and adhesins. *J Infect Dis,* 204**,** 704-13.

NOVICK, R. P. 2003. Mobile genetic elements and bacterial toxinoses: the superantigen-encoding pathogenicity islands of *Staphylococcus aureus*. *Plasmid,* 49**,** 93-105.

NOVICK, R. P., CHRISTIE, G. E. & PENADES, J. R. 2010. The phage-related chromosomal islands of Gram-positive bacteria. *Nat Rev Microbiol,* 8**,** 541-51.

NOVICK, R. P. & SUBEDI, A. 2007. The SaPIs: mobile pathogenicity islands of *Staphylococcus*. *Chem Immunol Allergy,* 93**,** 42-57.

NUBEL, U., DORDEL, J., KURT, K., STROMMENGER, B., WESTH, H., SHUKLA, S. K., ZEMLICKOVA, H., LEBLOIS, R., WIRTH, T., JOMBART, T., BALLOUX, F. & WITTE, W. 2010. A timescale for evolution, population expansion, and spatial spread of an emerging clone of methicillin-resistant *Staphylococcus aureus*. *PLoS Pathog,* 6**,** e1000855.

OKASHA, S. 2012. Population Genetics. *In:* ZALTA, E. N. (ed.) Fall 2012 ed.

OKORO, C. K., KINGSLEY, R. A., CONNOR, T. R., HARRIS, S. R., PARRY, C. M., AL-MASHHADANI, M. N., KARIUKI, S., MSEFULA, C. L., GORDON, M. A., DE PINNA, E., WAIN, J., HEYDERMAN, R. S.,

OBARO, S., ALONSO, P. L., MANDOMANDO, I., MACLENNAN, C. A., TAPIA, M. D., LEVINE, M. M., TENNANT, S. M., PARKHILL, J. & DOUGAN, G. 2012. Intracontinental spread of human invasive *Salmonella* Typhimurium pathovariants in sub-Saharan Africa. *Nat Genet,* 44**,** 1215-21.

OLIVER, A. 2010. Mutators in cystic fibrosis chronic lung infection: Prevalence, mechanisms, and consequences for antimicrobial therapy. *Int J Med Microbiol,* 300**,** 563-72.

OMRAN, A. R. 1971. The epidemiologic transition. A theory of the epidemiology of population change. *Milbank Mem Fund Q,* 49**,** 509-38.

ORRISON, L. H., CHERRY, W. B. & MILAN, D. 1981. Isolation of *Legionella pneumophilia* from cooling tower water by filtration. *Appl Environ Microbiol,* 41**,** 1202-5.

OSTERHOLM, M. T. & FORFANG, J. C. 1982. Toxic-shock syndrome in Minnesota: results of an active-passive surveillance system. *J Infect Dis,* 145**,** 458-64.

OU, C. Y., CIESIELSKI, C. A., MYERS, G., BANDEA, C. I., LUO, C. C., KORBER, B. T., MULLINS, J. I., SCHOCHETMAN, G., BERKELMAN, R. L., ECONOMOU, A. N. & ET AL. 1992. Molecular epidemiology of HIV transmission in a dental practice. *Science,* 256**,** 1165-71.

PAGANI, I., LIOLIOS, K., JANSSON, J., CHEN, I. M., SMIRNOVA, T., NOSRAT, B., MARKOWITZ, V. M. & KYRPIDES, N. C. 2012. The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res,* 40**,** D571-9.

PEACOCK, S. J., MOORE, C. E., JUSTICE, A., KANTZANOU, M., STORY, L., MACKIE, K., O'NEILL, G. & DAY, N. P. J. 2002. Virulent combinations of

adhesin and toxin genes in natural populations of *Staphylococcus aureus*. *Infect Immun,* 70**,** 4987-96.

PENTERMAN, J., NGUYEN, D., ANDERSON, E., STAUDINGER, B. J., GREENBERG, E. P., LAM, J. S. & SINGH, P. K. 2014. Rapid Evolution of Culture-Impaired Bacteria during Adaptation to Biofilm Growth. *Cell Rep*.

PEREZ-ROTH, E., LORENZO-DIAZ, F., BATISTA, N., MORENO, A. & MENDEZ-ALVAREZ, S. 2004. Tracking methicillin-resistant *Staphylococcus aureus* clones during a 5-year period (1998 to 2002) in a Spanish hospital. *J Clin Microbiol,* 42**,** 4649-56.

POHL, K., FRANCOIS, P., STENZ, L., SCHLINK, F., GEIGER, T., HERBERT, S., GOERKE, C., SCHRENZEL, J. & WOLZ, C. 2009. CodY in *Staphylococcus aureus*: a regulatory link between metabolism and virulence gene expression. *J Bacteriol,* 191**,** 2953-63.

POPOVICH, K. J., WEINSTEIN, R. A. & HOTA, B. 2008. Are community-associated methicillin-resistant *Staphylococcus aureus* (MRSA) strains replacing traditional nosocomial MRSA strains? *Clin Infect Dis,* 46**,** 787-94.

POTTS, A., SINKA, K., LOVE, J., DONAGHY, M., MCDONALD, E., JOHNSTON, F., BLATCHFORD, O., LINDSAY, D., BROWN, A. & EDWARDS, G. 2011. Legionellosis in Scotland in 2009 and 2010. *HPS Weekly Reports,* 45**,** 389-394.

POZZI, C., WATERS, E. M., RUDKIN, J. K., SCHAEFFER, C. R., LOHAN, A. J., TONG, P., LOFTUS, B. J., PIER, G. B., FEY, P. D., MASSEY, R. C. & O'GARA, J. P. 2012. Methicillin resistance alters the biofilm phenotype and attenuates virulence in *Staphylococcus aureus* device-associated infections. *PLoS Pathog,* 8**,** e1002626.

PRÉVOST, G., POTTECHER, B., DAHLET, M., BIENTZ, M., MANTZ, J. M. & PIÉMONT, Y. 1991. Pulsed field gel electrophoresis as a new epidemiological tool for monitoring methicillin-resistant *Staphylococcus aureus* in an intensive care unit. *J Hosp Infect,* 17**,** 255-69.

PRICE, L. B., STEGGER, M., HASMAN, H., AZIZ, M., LARSEN, J., ANDERSEN, P. S., PEARSON, T., WATERS, A. E., FOSTER, J. T., SCHUPP, J., GILLECE, J., DRIEBE, E., LIU, C. M., SPRINGER, B., ZDOVC, I., BATTISTI, A., FRANCO, A., ZMUDZKI, J., SCHWARZ, S., BUTAYE, P., JOUY, E., POMBA, C., PORRERO, M. C., RUIMY, R., SMITH, T. C., ROBINSON, D. A., WEESE, J. S., ARRIOLA, C. S., YU, F., LAURENT, F., KEIM, P., SKOV, R. & AARESTRUP, F. M. 2012. *Staphylococcus aureus* CC398: host adaptation and emergence of methicillin resistance in livestock. *MBio,* 3.

PRIEST, N. K., RUDKIN, J. K., FEIL, E. J., VAN DEN ELSEN, J. M., CHEUNG, A., PEACOCK, S. J., LAABEI, M., LUCKS, D. A., RECKER, M. & MASSEY, R. C. 2012. From genotype to phenotype: can systems biology be used to predict *Staphylococcus aureus* virulence? *Nat Rev Microbiol,* 10**,** 791-7.

PRUNIER, A. L. & LECLERCQ, R. 2005. Role of *mutS* and *mutL* genes in hypermutability and recombination in *Staphylococcus aureus*. *J Bacteriol,* 187**,** 3455-64.

PRUNIER, A. L., MALBRUNY, B., LAURANS, M., BROUARD, J., DUHAMEL, J. F. & LECLERCQ, R. 2003. High rate of macrolide resistance in *Staphylococcus aureus* strains from patients with cystic fibrosis reveals high proportions of hypermutable strains. *J Infect Dis,* 187**,** 1709-16.

QUECK, S. Y., KHAN, B. A., WANG, R., BACH, T. H., KRETSCHMER, D., CHEN, L., KREISWIRTH, B. N., PESCHEL, A., DELEO, F. R. & OTTO,

M. 2009. Mobile genetic element-encoded cytolysin connects virulence to methicillin resistance in MRSA. *PLoS Pathog,* 5**,** e1000533.

RASKO, D. A., WEBSTER, D. R., SAHL, J. W., BASHIR, A., BOISEN, N., SCHEUTZ, F., PAXINOS, E. E., SEBRA, R., CHIN, C. S., ILIOPOULOS, D., KLAMMER, A., PELUSO, P., LEE, L., KISLYUK, A. O., BULLARD, J., KASARSKIS, A., WANG, S., EID, J., RANK, D., REDMAN, J. C., STEYERT, S. R., FRIMODT-MOLLER, J., STRUVE, C., PETERSEN, A. M., KROGFELT, K. A., NATARO, J. P., SCHADT, E. E. & WALDOR, M. K. 2011. Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany. *N Engl J Med,* 365**,** 709-17.

RATZOW, S., GAIA, V., HELBIG, J. H., FRY, N. K. & LUCK, P. C. 2007. Addition of *neuA,* the gene encoding N-acylneuraminate cytidylyl transferase, increases the discriminatory ability of the consensus sequence-based scheme for typing *Legionella pneumophila* serogroup 1 strains. *J Clin Microbiol,* 45**,** 1965-8.

RECSEI, P., KREISWIRTH, B., O'REILLY, M., SCHLIEVERT, P., GRUSS, A. & NOVICK, R. P. 1986. Regulation of exoprotein gene expression in *Staphylococcus aureus* by *agr*. *MGG Molecular & General Genetics,* 202**,** 58-61.

REGNERY, R. L., ANDERSON, B. E., CLARRIDGE, J. E., 3RD, RODRIGUEZ-BARRADAS, M. C., JONES, D. C. & CARR, J. H. 1992. Characterization of a novel *Rochalimaea* species, *R. henselae* sp. nov., isolated from blood of a febrile, human immunodeficiency virus-positive patient. *J Clin Microbiol,* 30**,** 265-74.

REUTER, S., HARRISON, T. G., KOSER, C. U., ELLINGTON, M. J., SMITH, G. P., PARKHILL, J., PEACOCK, S. J., BENTLEY, S. D. & TOROK, M. E. 2013. A pilot study of rapid whole-genome sequencing for the investigation

of a *Legionella* outbreak. *BMJ Open*, 3, e002175.

RIDDER-SCHAPHORN, S., RATJEN, F., DUBBERS, A., HABERLE, J., FALK, S., KUSTER, P., SCHUSTER, A., MELLIES, U., LOWE, B., REINTJES, R., PETERS, G. & KAHL, B. C. 2007. Nasal *Staphylococcus aureus* carriage is not a risk factor for lower-airway infection in young cystic fibrosis patients. *J Clin Microbiol,* 45**,** 2979-84.

RIDENOUR, D. A., CIRILLO, S. L. G., FENG, S., SAMRAKANDI, M. M. & CIRILLO, J. D. 2003. Identification of a gene that affects the efficiency of host cell infection by *Legionella pneumophila* in a temperature-dependent fashion. *Infect Immun,* 71**,** 6256-63.

RILEY, L. W., REMIS, R. S., HELGERSON, S. D., MCGEE, H. B., WELLS, J. G., DAVIS, B. R., HEBERT, R. J., OLCOTT, E. S., JOHNSON, L. M., HARGRETT, N. T., BLAKE, P. A. & COHEN, M. L. 1983. Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. *N Engl J Med,* 308**,** 681-5.

RIORDAN, J., ROMMENS, J., KEREM, B., ALON, N., ROZMAHEL, R., GRZELCZAK, Z., ZIELENSKI, J., LOK, S., PLAVSIC, N., CHOU, J., DRUMM, M., IANNUZZI, M., COLLINS, F. & TSUI, L. 1989. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science,* 245**,** 1066-1073.

ROBINSON, D. A. & ENRIGHT, M. C. 2004. Evolution of *Staphylococcus aureus* by large chromosomal replacements. *J Bacteriol,* 186**,** 1060-4.

ROBINSON, D. A., KEARNS, A. M., HOLMES, A., MORRISON, D., GRUNDMANN, H., EDWARDS, G., O'BRIEN, F. G., TENOVER, F. C., MCDOUGAL, L. K., MONK, A. B. & ENRIGHT, M. C. 2005. Re-emergence of early pandemic *Staphylococcus aureus* as a community-

acquired meticillin-resistant clone. *Lancet,* 365**,** 1256-8.

ROBINSON, D. A., MONK, A. B., COOPER, J. E., FEIL, E. J. & ENRIGHT, M. C. 2005. Evolutionary genetics of the accessory gene regulator (*agr*) locus in *Staphylococcus aureus*. *J Bacteriol,* 187**,** 8312-21.

ROCHA, E. P., SMITH, J. M., HURST, L. D., HOLDEN, M. T., COOPER, J. E., SMITH, N. H. & FEIL, E. J. 2006. Comparisons of dN/dS are time dependent for closely related bacterial genomes. *J Theor Biol,* 239**,** 226-35.

ROGERS, J., DOWSETT, A. B., DENNIS, P. J., LEE, J. V. & KEEVIL, C. W. 1994. Influence of temperature and plumbing material selection on biofilm formation and growth of *Legionella pneumophila* in a model potable water system containing complex microbial flora. *Appl Environ Microbiol,* 60**,** 1585-92.

ROHDE, H., QIN, J., CUI, Y., LI, D., LOMAN, N. J., HENTSCHKE, M., CHEN, W., PU, F., PENG, Y., LI, J., XI, F., LI, S., LI, Y., ZHANG, Z., YANG, X., ZHAO, M., WANG, P., GUAN, Y., CEN, Z., ZHAO, X., CHRISTNER, M., KOBBE, R., LOOS, S., OH, J., YANG, L., DANCHIN, A., GAO, G. F., SONG, Y., LI, Y., YANG, H., WANG, J., XU, J., PALLEN, M. J., WANG, J., AEPFELBACHER, M., YANG, R. & CONSORTIUM, E. C. O. H. G. A. C.-S. 2011. Open-source genomic analysis of Shiga-toxin-producing *E. coli* O104:H4. *N Engl J Med,* 365**,** 718-24.

ROMMENS, J. M., IANNUZZI, M. C., KEREM, B., DRUMM, M. L., MELMER, G., DEAN, M., ROZMAHEL, R., COLE, J. L., KENNEDY, D., HIDAKA, N. & ET AL. 1989. Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science,* 245**,** 1059-65.

ROSDAHL, V. T. & KNUDSEN, A. M. 1991. The decline of methicillin resistance among Danish *Staphylococcus aureus* strains. *Infect Control Hosp Epidemiol,*

12**,** 83-8.

ROSENDAL, K., JESSEN, O., BENTZON, M. W. & BÜLOW, P. 1977. Antibiotic policy and spread of *Staphylococcus aureus* strains in Danish hospitals, 1969-1974. *Acta pathologica et microbiologica Scandinavica. Section B, Microbiology,* 85**,** 143-52.

ROSSNEY, A. S., SHORE, A. C., MORGAN, P. M., FITZGIBBON, M. M., O'CONNELL, B. & COLEMAN, D. C. 2007. The emergence and importation of diverse genotypes of methicillin-resistant *Staphylococcus aureus* (MRSA) harboring the Panton-Valentine leukocidin gene (*pvl*) reveal that *pvl* is a poor marker for community-acquired MRSA strains in Ireland. *J Clin Microbiol,* 45**,** 2554-63.

ROUNTREE, P. M. & ASHESHOV, E. H. 1961. Further Observations on Changes in the Phage-Typing Pattern of Phage Type 80/81 Staphylococci. *J Gen Microbiol,* 26**,** 111-122.

ROWBOTHAM, T. J. 1980. Preliminary report on the pathogenicity of *Legionella pneumophila* for freshwater and soil amoebae. *J Clin Microbiol,* 33**,** 1179-83.

RUDBECK, M., MOLBAK, K. & ULDUM, S. A. 2009. Dynamics of *Legionella* antibody levels during 1 year in a healthy population. *Epidemiol Infect,* 137**,** 1013-8.

RUDKIN, J. K., EDWARDS, A. M., BOWDEN, M. G., BROWN, E. L., POZZI, C., WATERS, E. M., CHAN, W. C., WILLIAMS, P., O'GARA, J. P. & MASSEY, R. C. 2012. Methicillin resistance reduces the virulence of healthcare-associated methicillin-resistant *Staphylococcus aureus* by interfering with the *agr* quorum sensing system. *J Infect Dis,* 205**,** 798-806.

SADOSKY, A. B., WIATER, L. A. & SHUMAN, H. A. 1993. Identification of

Legionella pneumophila genes required for growth within and killing of human macrophages. *Infect Immun,* 61**,** 5361-73.

SAMPAIO, J. L. M., VIANA-NIERO, C., DE FREITAS, D., HÖFLING-LIMA, A. L. & LEÃO, S. C. 2006. Enterobacterial repetitive intergenic consensus PCR is a useful tool for typing *Mycobacterium chelonae* and *Mycobacterium abscessus* isolates. *Diagn Microbiol Infect Dis.,* 55**,** 107-18.

SANCHEZ, I., GARCIA-NUNEZ, M., RAGULL, S., SOPENA, N., PEDRO-BOTET, M. L., REY-JOLY, C., SABRIA, M. & ESTEVE, M. 2008. Genotypic variability and persistence of *Legionella pneumophila* PFGE patterns in 34 cooling towers from two different areas. *Environ Microbiol,* 10**,** 395-9.

SASS, P., BERSCHEID, A., JANSEN, A., OEDENKOVEN, M., SZEKAT, C., STRITTMATTER, A., GOTTSCHALK, G. & BIERBAUM, G. 2012. Genome sequence of *Staphylococcus aureus* VC40, a vancomycin- and daptomycin-resistant strain, to study the genetics of development of resistance to currently applied last-resort antibiotics. *J Bacteriol,* 194**,** 2107-8.

SCHLECH, W. F., 3RD, LAVIGNE, P. M., BORTOLUSSI, R. A., ALLEN, A. C., HALDANE, E. V., WORT, A. J., HIGHTOWER, A. W., JOHNSON, S. E., KING, S. H., NICHOLLS, E. S. & BROOME, C. V. 1983. Epidemic listeriosis--evidence for transmission by food. *N Engl J Med,* 308**,** 203-6.

SCHOULS, L. M., SPALBURG, E. C., VAN LUIT, M., HUIJSDENS, X. W., PLUISTER, G. N., VAN SANTEN-VERHEUVEL, M. G., VAN DER HEIDE, H. G., GRUNDMANN, H., HECK, M. E. & DE NEELING, A. J. 2009. Multiple-locus variable number tandem repeat analysis of *Staphylococcus aureus*: comparison with pulsed-field gel electrophoresis and spa-typing. *PLoS One,* 4**,** e5082.

SCHROEDER, G. N., PETTY, N. K., MOUSNIER, A., HARDING, C. R., VOGRIN, A. J., WEE, B., FRY, N. K., HARRISON, T. G., NEWTON, H. J., THOMSON, N. R., BEATSON, S. A., DOUGAN, G., HARTLAND, E. L. & FRANKEL, G. 2010. *Legionella pneumophila* strain 130b possesses a unique combination of type IV secretion systems and novel Dot/Icm secretion system effector proteins. *J Bacteriol,* 192**,** 6001-16.

SCHULTE, B., BIERBAUM, G., POHL, K., GOERKE, C. & WOLZ, C. 2013. Diversification of clonal complex 5 methicillin-resistant *Staphylococcus aureus* strains (Rhine-Hesse clone) within Germany. *J Clin Microbiol,* 51**,** 212-6.

SCICLUNA, E. A., SHORE, A. C., THURMER, A., EHRICHT, R., SLICKERS, P., BORG, M. A., COLEMAN, D. C. & MONECKE, S. 2010. Characterisation of MRSA from Malta and the description of a Maltese epidemic MRSA strain. *Eur J Clin Microbiol Infect Dis,* 29**,** 163-70.

SEIDL, K., BISCHOFF, M. & BERGER-BACHI, B. 2008. CcpA mediates the catabolite repression of *tst* in *Staphylococcus aureus*. *Infect Immun,* 76**,** 5093-9.

SEIDL, K., STUCKI, M., RUEGG, M., GOERKE, C., WOLZ, C., HARRIS, L., BERGER-BACHI, B. & BISCHOFF, M. 2006. *Staphylococcus aureus* CcpA affects virulence determinant production and antibiotic resistance. *Antimicrob Agents Chemother,* 50**,** 1183-94.

SETH-SMITH, H. M., HARRIS, S. R., SKILTON, R. J., RADEBE, F. M., GOLPARIAN, D., SHIPITSYNA, E., DUY, P. T., SCOTT, P., CUTCLIFFE, L. T., O'NEILL, C., PARMAR, S., PITT, R., BAKER, S., ISON, C. A., MARSH, P., JALAL, H., LEWIS, D. A., UNEMO, M., CLARKE, I. N., PARKHILL, J. & THOMSON, N. R. 2013. Whole-genome sequences of *Chlamydia trachomatis* directly from clinical samples without culture.

Genome Res, 23**,** 855-66.

SHANDS, K. N., SCHMID, G. P., DAN, B. B., BLUM, D., GUIDOTTI, R. J., HARGRETT, N. T., ANDERSON, R. L., HILL, D. L., BROOME, C. V., BAND, J. D. & FRASER, D. W. 1980. Toxic-shock syndrome in menstruating women: association with tampon use and *Staphylococcus aureus* and clinical features in 52 cases. *N Engl J Med,* 303**,** 1436-42.

SHERWOOD, J., PARK, M., ROBBEN, P., WHITMAN, T. & ELLIS, M. W. 2013. USA300 methicillin-resistant *Staphylococcus aureus* emerging as a cause of bloodstream infections at military medical centers. *Infect Control Hosp Epidemiol,* 34**,** 393-9.

SIEVERT, D. M., RUDRIK, J. T., PATEL, J. B., MCDONALD, L. C., WILKINS, M. J. & HAGEMAN, J. C. 2008. Vancomycin-resistant *Staphylococcus aureus* in the United States, 2002-2006. *Clin Infect Dis,* 46**,** 668-74.

SMITH, E. E., BUCKLEY, D. G., WU, Z., SAENPHIMMACHAK, C., HOFFMAN, L. R., D'ARGENIO, D. A., MILLER, S. I., RAMSEY, B. W., SPEERT, D. P., MOSKOWITZ, S. M., BURNS, J. L., KAUL, R. & OLSON, M. V. 2006. Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. *Proc Natl Acad Sci U S A,* 103**,** 8487-92.

SMYTH, E. T., MCILVENNY, G., ENSTONE, J. E., EMMERSON, A. M., HUMPHREYS, H., FITZPATRICK, F., DAVIES, E., NEWCOMBE, R. G., SPENCER, R. C. & HOSPITAL INFECTION SOCIETY PREVALENCE SURVEY STEERING, G. 2008. Four country healthcare associated infection prevalence survey 2006: overview of the results. *J Hosp Infect,* 69**,** 230-48.

SODERQUIST, B., BERGLUND, C. & STRALIN, K. 2006. Community-acquired pneumonia and bacteremia caused by an unusual methicillin-resistant *Staphylococcus aureus* (MRSA) strain with sequence type 36, staphylococcal

cassette chromosome *mec* type IV and Panton-Valentine leukocidin genes. *Eur J Clin Microbiol Infect Dis,* 25**,** 604-6.

SOUZA, H. A., NOGUEIRA, K. S., MATOS, A. P., VIEIRA, R. P., RIEDI, C. A., ROSARIO, N. A., TELLES, F. Q. & COSTA, L. M. 2006. Early microbial colonization of cystic fibrosis patients identified by neonatal screening, with emphasis on *Staphylococcus aureus*. *J Pediatr (Rio J),* 82**,** 377-82.

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

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. *Int J Antimicrob Agents,* 39**,** 273-82.

STEVENS, D. L., TANNER, M. H., WINSHIP, J., SWARTS, R., RIES, K. M., SCHLIEVERT, P. M. & KAPLAN, E. 1989. Severe group A streptococcal infections associated with a toxic shock-like syndrome and scarlet fever toxin A. *N Engl J Med,* 321**,** 1-7.

STRYJEWSKI, M. E. & CHAMBERS, H. F. 2008. Skin and soft-tissue infections caused by community-acquired methicillin-resistant *Staphylococcus aureus*. *Clin Infect Dis,* 46 Suppl 5**,** S368-77.

SUBEDI, A., UBEDA, C., ADHIKARI, R. P., PENADES, J. R. & NOVICK, R. P. 2007. Sequence analysis reveals genetic exchanges and intraspecific spread of SaPI2, a pathogenicity island involved in menstrual toxic shock. *Microbiology,* 153**,** 3235-45.

TANAKA, M., WANG, T., ONODERA, Y., UCHIDA, Y. & SATO, K. 2000.

Mechanism of quinolone resistance in *Staphylococcus aureus*. *J Infect Chemother,* 6**,** 131-9.

TATEDA, K., DENG, J. C., MOORE, T. A., NEWSTEAD, M. W., PAINE, R., 3RD, KOBAYASHI, N., YAMAGUCHI, K. & STANDIFORD, T. J. 2003. Hyperoxia mediates acute lung injury and increased lethality in murine *Legionella pneumonia*: the role of apoptosis. *J Immunol,* 170**,** 4209-16.

TENAILLON, O., SKURNIK, D., PICARD, B. & DENAMUR, E. 2010. The population genetics of commensal *Escherichia coli*. *Nat Rev Microbiol,* 8**,** 207-17.

THOMAS, D. L., MUNDY, L. M. & TUCKER, P. C. 1993. Hot tub legionellosis. Legionnaires' disease and Pontiac fever after a point-source exposure to *Legionella pneumophila*. *Arch Intern Med,* 153**,** 2597-9.

THOMAS, J. C., GODFREY, P. A., FELDGARDEN, M. & ROBINSON, D. A. 2012. Candidate targets of balancing selection in the genome of *Staphylococcus aureus*. *Mol Biol Evol,* 29**,** 1175-86.

THUMANU, K., CHA, J., FISHER, J. F., PERRINS, R., MOBASHERY, S. & WHARTON, C. 2006. Discrete steps in sensing of beta-lactam antibiotics by the BlaR1 protein of the methicillin-resistant *Staphylococcus aureus* bacterium. *Proc Natl Acad Sci U S A,* 103**,** 10630-5.

TIEMERSMA, E. W., BRONZWAER, S. L. A. M., LYYTIKÄINEN, O., DEGENER, J. E., SCHRIJNEMAKERS, P., BRUINSMA, N., MONEN, J., WITTE, W. & GRUNDMAN, H. 2004. Methicillin-resistant *Staphylococcus aureus* in Europe, 1999-2002. *Emerg Infect Dis,* 10**,** 1627-34.

TONG, S. Y., SHARMA-KUINKEL, B. K., THADEN, J. T., WHITNEY, A. R., YANG, S. J., MISHRA, N. N., RUDE, T., LILLIEBRIDGE, R. A., SELIM,

M. A., AHN, S. H., HOLT, D. C., GIFFARD, P. M., BAYER, A. S., DELEO, F. R. & FOWLER, V. G., JR. 2013. Virulence of endemic nonpigmented northern Australian *Staphylococcus aureus* clone (clonal complex 75, *S. argenteus*) is not augmented by staphyloxanthin. *J Infect Dis,* 208**,** 520-7.

TORMO-MAS, M. A., MIR, I., SHRESTHA, A., TALLENT, S. M., CAMPOY, S., LASA, I., BARBE, J., NOVICK, R. P., CHRISTIE, G. E. & PENADES, J. R. 2010. Moonlighting bacteriophage proteins derepress staphylococcal pathogenicity islands. *Nature,* 465**,** 779-82.

TORRES-SANGIAO, E., PEREZ-CASTRO, S., FERNANDEZ-NATAL, M. I., CISTERNA-CANCER, R., ZAPICO-GONZALEZ, M., FERNANDEZ-PEREZ, B., OJEDA-FERNANDEZ, E., NEBREDA, T., GOZALO-MARGUELLO, M., FUSTER-FOZ, C., ROIZ-MESONES, M. P., MIGUEL-MARTIN, M. D., TORROBA, L., COIRA-NIETO, A., VASALLO-VIDAL, F., MENDEZ-LAGE, S., PRIETO-RODRIGUEZ, E., EIROS, J. M., TORRES, J. & BOU, G. 2012. Identification of international circulating lineages of meticillin-resistant *Staphylococcus aureus* in the north of Spain and their glycopeptide and linezolid susceptibility. *J Med Microbiol,* 61**,** 305-7.

TUCHSCHERR, L., MEDINA, E., HUSSAIN, M., VOLKER, W., HEITMANN, V., NIEMANN, S., HOLZINGER, D., ROTH, J., PROCTOR, R. A., BECKER, K., PETERS, G. & LOFFLER, B. 2011. *Staphylococcus aureus* phenotype switching: an effective bacterial strategy to escape host immune response and establish a chronic infection. *EMBO Mol Med,* 3**,** 129-41.

UBUKATA, K., NONOGUCHI, R., MATSUHASHI, M. & KONNO, M. 1989. Expression and inducibility in *Staphylococcus aureus* of the *mecA* gene, which encodes a methicillin-resistant *S. aureus* - specific penicillin-binding protein. *J Bacteriol,* 171**,** 2882-5.

VAN BELKUM, A., VAN LEEUWEN, W., KAUFMANN, M. E., COOKSON, B., FOREY, F., ETIENNE, J., GOERING, R., TENOVER, F., STEWARD, C., O'BRIEN, F., GRUBB, W., TASSIOS, P., LEGAKIS, N., MORVAN, A., EL SOLH, N., DE RYCK, R., STRUELENS, M., SALMENLINNA, S., VUOPIO-VARKILA, J., KOOISTRA, M., TALENS, A., WITTE, W. & VERBRUGH, H. 1998. Assessment of resolution and intercenter reproducibility of results of genotyping *Staphylococcus aureus* by pulsed-field gel electrophoresis of SmaI macrorestriction fragments: a multicenter study. *J Clin Microbiol,* 36**,** 1653-9.

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**,** 1820-6.

VAN WAMEL, W. J., ROOIJAKKERS, S. H., RUYKEN, M., VAN KESSEL, K. P. & VAN STRIJP, J. A. 2006. The innate immune modulators staphylococcal complement inhibitor and chemotaxis inhibitory protein of *Staphylococcus aureus* are located on beta-hemolysin-converting bacteriophages. *J Bacteriol,* 188**,** 1310-5.

VICKERS, A. A., POTTER, N. J., FISHWICK, C. W., CHOPRA, I. & O'NEILL, A. J. 2009. Analysis of mutational resistance to trimethoprim in *Staphylococcus aureus* by genetic and structural modelling techniques. *J Antimicrob Chemother,* 63**,** 1112-7.

VU-THIEN, H., HORMIGOS, K., CORBINEAU, G., FAUROUX, B., CORVOL, H., MOISSENET, D., VERGNAUD, G. & POURCEL, C. 2010. Longitudinal survey of *Staphylococcus aureus* in cystic fibrosis patients using a multiple-locus variable-number of tandem-repeats analysis method. *BMC Microbiol,* 10**,** 24.

WARD, M., BOLAND, M., NICOLAY, N., MURPHY, H., MCELHINEY, J., COLLINS, C., LYNCH, M., MCCARTHY, M. & J, O. D. 2010. A cluster of Legionnaires' disease and associated Pontiac fever morbidity in office workers, Dublin, June-July 2008. *J Environ Public Health,* 2010**,** 463926.

WATANABE, S., ITO, T., SASAKI, T., LI, S., UCHIYAMA, I., KISHII, K., KIKUCHI, K., SKOV, R. L. & HIRAMATSU, K. 2009. Genetic diversity of staphylocoagulase genes (*coa*): insight into the evolution of variable chromosomal virulence factors in *Staphylococcus aureus*. *PLoS One,* 4**,** e5714.

WATTS, D. J. & STROGATZ, S. H. 1998. Collective dynamics of 'small-world' networks. *Nature,* 393**,** 440-2.

WERTHEIM, H., MELLES, D., VOS, M., VANLEEUWEN, W., VANBELKUM, A., VERBRUGH, H. & NOUWEN, J. 2005. The role of nasal carriage in infections. *Lancet Infect Dis,* 5**,** 751-762.

WERTHEIM, H. F., VOS, M. C., OTT, A., VAN BELKUM, A., VOSS, A., KLUYTMANS, J. A., VAN KEULEN, P. H., VANDENBROUCKE-GRAULS, C. M., MEESTER, M. H. & VERBRUGH, H. A. 2004. Risk and outcome of nosocomial *Staphylococcus aureus* bacteraemia in nasal carriers versus non-carriers. *Lancet,* 364**,** 703-5.

WHITE, P. S., GRAHAM, F. F., HARTE, D. J., BAKER, M. G., AMBROSE, C. D. & HUMPHREY, A. R. 2013. Epidemiological investigation of a Legionnaires' disease outbreak in Christchurch, New Zealand: the value of spatial methods for practical public health. *Epidemiol Infect,* 141**,** 789-99.

WICHELHAUS, T. A., BODDINGHAUS, B., BESIER, S., SCHAFER, V., BRADE, V. & LUDWIG, A. 2002. Biological cost of rifampin resistance from the perspective of *Staphylococcus aureus*. *Antimicrob Agents*

Chemother, 46**,** 3381-5.

WILGENBUSCH, J. C. & SWOFFORD, D. 2003. Inferring evolutionary trees with PAUP*. *Curr Protoc Bioinformatics,* Chapter 6**,** Unit 6 4.

WILSON, G. J., SEO, K. S., CARTWRIGHT, R. A., CONNELLEY, T., CHUANG-SMITH, O. N., MERRIMAN, J. A., GUINANE, C. M., PARK, J. Y., BOHACH, G. A., SCHLIEVERT, P. M., MORRISON, W. I. & FITZGERALD, J. R. 2011. A novel core genome-encoded superantigen contributes to lethality of community-associated MRSA necrotizing pneumonia. *PLoS Pathog,* 7**,** e1002271.

WIRTH, T., WANG, X., LINZ, B., NOVICK, R. P., LUM, J. K., BLASER, M., MORELLI, G., FALUSH, D. & ACHTMAN, M. 2004. Distinguishing human ethnic groups by means of sequences from *Helicobacter pylori*: lessons from Ladakh. *Proc Natl Acad Sci U S A,* 101**,** 4746-51.

WONG, A. & KASSEN, R. 2011. Parallel evolution and local differentiation in quinolone resistance in *Pseudomonas aeruginosa*. *Microbiology,* 157**,** 937-44.

WYLLIE, D. H., WALKER, A. S., MILLER, R., MOORE, C., WILLIAMSON, S. R., SCHLACKOW, I., FINNEY, J. M., O'CONNOR, L., PETO, T. E. & CROOK, D. W. 2011. Decline of meticillin-resistant *Staphylococcus aureus* in Oxfordshire hospitals is strain-specific and preceded infection-control intensification. *BMJ Open,* 1**,** e000160.

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**,** 4550-5.

ZDZIARSKI, J., BRZUSZKIEWICZ, E., WULLT, B., LIESEGANG, H., BIRAN, D., VOIGT, B., GRONBERG-HERNANDEZ, J., RAGNARSDOTTIR, B., HECKER, M., RON, E. Z., DANIEL, R., GOTTSCHALK, G., HACKER, J., SVANBORG, C. & DOBRINDT, U. 2010. Host imprints on bacterial genomes--rapid, divergent evolution in individual patients. *PLoS Pathog,* 6**,** e1001078.

ZERBINO, D. R. & BIRNEY, E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res,* 18**,** 821-9.

ZHANG, K., MCCLURE, J. A., ELSAYED, S., LOUIE, T. & CONLY, J. M. 2005. Novel multiplex PCR assay for characterization and concomitant subtyping of staphylococcal cassette chromosome *mec* types I to V in methicillin-resistant *Staphylococcus aureus*. *J Clin Microbiol,* 43**,** 5026-33.

ZUSMAN, T., YERUSHALMI, G. & SEGAL, G. 2003. Functional similarities between the *icm*/*dot* pathogenesis systems of *Coxiella burnetii* and *Legionella pneumophila*. *Infect Immun,* 71**,** 3714-3723.

Appendix 1

Details of CC30 isolates used in this study

Isolate	ST	Lineage	Isolation Year	Country	Reference
00_4016_R	30	-	2000	Scotland	This study
02_7796_N	30	-	2002	Scotland	This study
05T116	30	-	2005	Sweden	Söderquist et al., 2006
09_7338_H	30	-	2009	Scotland	This study
4022	30	-	1998	Cuba	Robinson et al., 2005
MSA3412	37	-	1980s	USA	Fitzgerald et al., 2001
MSA2885	441	-	1974	Canada	Fitzgerald et al., 2001
55.2053_55	30	80/81	1955	England	Robinson et al., 2005
55.2798_55	30	80/81	1955	England	Robinson et al., 2005
55.3224_55	30	80/81	1955	England	Robinson et al., 2005
58.362_58	30	80/81	1958	USA	Robinson et al., 2005
58.424_58	30	80/81	1958	USA	Robinson et al., 2005
65.1322_65	30	80/81	1965	USA	Robinson et al., 2005
65.20_65	30	80/81	1965	USA	Robinson et al., 2005
66.1888_66	30	80/81	1966	USA	Robinson et al., 2005
68.397_68	30	80/81	1968	USA	Robinson et al., 2005
68.81_68	30	80/81	1968	USA	Robinson et al., 2005

Isolate	ST	Lineage	Isolation Year	Country	Reference
69.172_69	30	80/81	1969	USA	Robinson et al., 2005
69.412_69	30	80/81	1969	USA	Robinson et al., 2005
A017934_97	30	80/81	1997	Sweden	Robinson et al., 2005
e1410_62	30	80/81	1962	Denmark	Robinson et al., 2005
M1015_62	30	80/81	1962	Australia	Robinson et al., 2005
M1016_62	30	80/81	1962	Australia	Robinson et al., 2005
M876_61	30	80/81	1961	Australia	Robinson et al., 2005
M899_61	30	80/81	1961	Australia	Robinson et al., 2005
1227_07.08	36	EMRSA-16	2007	Scotland	This study
1696_07.08	36	EMRSA-16	2007	Scotland	This study
1712_07.08	36	EMRSA-16	2007	Scotland	This study
1791_03.08	36	EMRSA-16	2003	Scotland	This study
2449_07.17	36	EMRSA-16	2007	Scotland	This study
2483_97.75	36	EMRSA-16	1997	Scotland	This study
2496_07.17	36	EMRSA-16	2007	Scotland	This study
2589_07.17	36	EMRSA-16	2007	Scotland	This study
2880_07.25	36	EMRSA-16	2007	Scotland	This study
3366_99	36	EMRSA-16	1999	Ireland	Shore et al., 2005

Isolate	ST	Lineage	Isolation Year	Country	Reference
3841_07.33	36	EMRSA-16	2007	Scotland	This study
5806_98.83	36	EMRSA-16	1998	Scotland	This study
6636_07.67	36	EMRSA-16	2007	Scotland	This study
6659_07.67	36	EMRSA-16	2007	Scotland	This study
7206_07.75	36	EMRSA-16	2007	Scotland	This study
75916_95	36	EMRSA-16	1995	Finland	Robinson et al., 2005
9570_06.92	36	EMRSA-16	2006	Scotland	This study
DEN4415_01	36	EMRSA-16	2001	Denmark	Faria et al., 2005
HAR24_93	36	EMRSA-16	1993	UK	Harmony Collection
HU275_02	36	EMRSA-16	2002	Hungary	Conceicao et al., 2007
MRSA252_97	36	EMRSA-16	1997	UK	Holden et al., 2004
sa1_98.92	36	EMRSA-16	1998	England	This study
sa10_07.75	36	EMRSA-16	2007	Scotland	This study
sa11_07.42	36	EMRSA-16	2007	Scotland	This study
sa12_07.67	36	EMRSA-16	2007	Scotland	This study
sa13_07.33	36	EMRSA-16	2007	Scotland	This study
sa14_06.50	36	EMRSA-16	2006	Scotland	This study
sa15_08.25	36	EMRSA-16	2008	England	This study

Isolate	ST	Lineage	Isolation Year	Country	Reference
sa16_06.75	36	EMRSA-16	2006	England	This study
sa17_06.50	36	EMRSA-16	2006	Scotland	This study
sa18_07.50	36	EMRSA-16	2007	England	This study
sa19_06.00	36	EMRSA-16	2005	England	This study
sa2_97.00	36	EMRSA-16	1997	England	This study
sa20_07.83	36	EMRSA-16	2007	England	This study
sa21_03.00	36	EMRSA-16	2003	England	This study
sa22_07.42	36	EMRSA-16	2007	England	This study
sa23_07.50	36	EMRSA-16	2007	England	This study
sa24_07.50	36	EMRSA-16	2007	England	This study
sa25_99.67	36	EMRSA-16	1999	England	This study
sa26_00.25	36	EMRSA-16	2000	England	This study
sa27_00.33	36	EMRSA-16	2000	England	This study
sa28_02.42	36	EMRSA-16	2002	England	This study
sa29_99.17	36	EMRSA-16	1999	England	This study
sa3_97.92	36	EMRSA-16	1997	England	This study
sa30_07.75	36	EMRSA-16	2007	England	This study
sa31_02.00	36	EMRSA-16	2002	England	This study

Isolate	ST	Lineage	Isolation Year	Country	Reference
sa32_07.42	36	EMRSA-16	2007	Scotland	This study
sa33_98.92	36	EMRSA-16	1998	England	This study
sa34_06.83	36	EMRSA-16	2006	Scotland	This study
sa35_07.50	36	EMRSA-16	2007	Scotland	This study
sa36_06.25	36	EMRSA-16	2006	Scotland	This study
sa37_07.33	36	EMRSA-16	2007	England	This study
sa38_05.83	36	EMRSA-16	2005	Scotland	This study
sa4_97.67	36	EMRSA-16	1997	England	This study
sa5_95.00	36	EMRSA-16	1995	Finland	This study
sa6_93.00	36	EMRSA-16	1993	England	This study
sa7_02.00	36	EMRSA-16	2002	England	This study
sa8_02.00	36	EMRSA-16	2002	England	This study
sa9_07.67	36	EMRSA-16	2007	Scotland	This study
7G_02	500	EMRSA-16	2002	USA	This study
PS12	500	EMRSA-16	-	USA	This study
MSA3407	36	EMRSA16	1980s	USA	Fitzgerald et al., 2001
02_7346_J	30	Other epidemic CC30	2002	Scotland	This study
04_2707_R	30	Other epidemic CC30	2004	Scotland	This study

Isolate	ST	Lineage	Isolation Year	Country	Reference
05_6324_F	30	Other epidemic CC30	2005	Scotland	This study
06_5394_X	30	Other epidemic CC30	2006	Scotland	This study
08_8403_F	30	Other epidemic CC30	2008	Scotland	This study
09_6122_D	30	Other epidemic CC30	2009	Scotland	This study
10_1316_Z	30	Other epidemic CC30	2010	Scotland	This study
10_6346_K	30	Other epidemic CC30	2010	Scotland	This study
11_3148_R	30	Other epidemic CC30	2011	Scotland	This study
11_3669_V	30	Other epidemic CC30	2011	Scotland	This study
11_4342	30	Other epidemic CC30	2011	Scotland	This study
11_4425	30	Other epidemic CC30	2011	Scotland	This study
11_5419_S	30	Other epidemic CC30	2011	Scotland	This study
4030	30	Other epidemic CC30	-	Cuba	Robinson et al., 2005
BTN1260	30	Other epidemic CC30	1999	England	Robinson et al., 2005
C101	30	Other epidemic CC30	1997	England	Robinson et al., 2005
ED83_83.92	30	Other epidemic CC30	1983	France	Branger et al., 1996
ED84_84.83	30	Other epidemic CC30	1984	France	Branger et al., 1996
ED86_86.08	30	Other epidemic CC30	1986	France	Branger et al., 1996
MN8_80	30	Other epidemic CC30	1980	USA	Altemeier et al., 1982

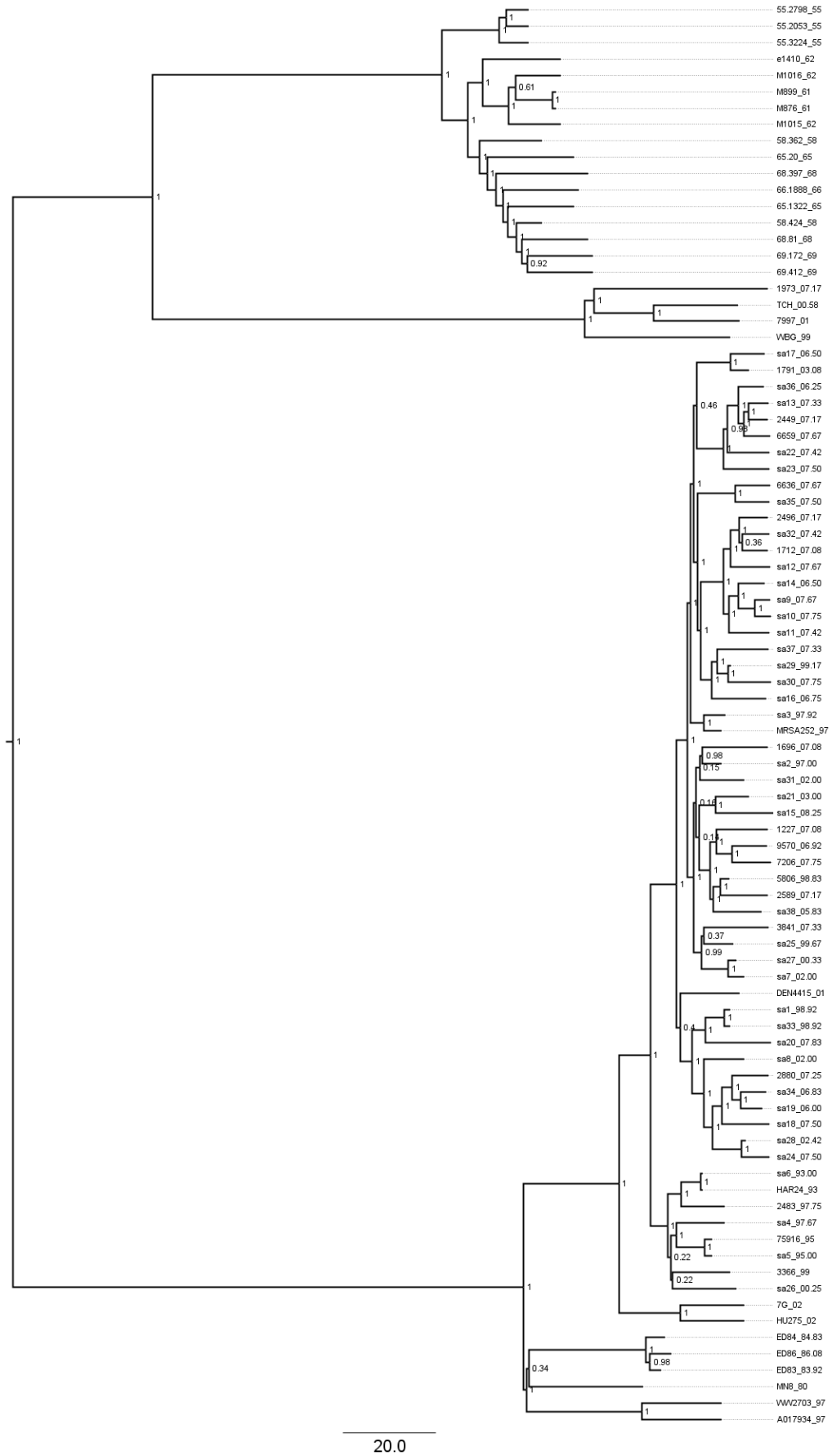
Isolate	ST	Lineage	Isolation Year	Country	Reference
MSA1205	30	Other epidemic CC30	-	USA	Fitzgerald et al., 2001
MSA1832	30	Other epidemic CC30	1968	USA	Fitzgerald et al., 2001
MSA2346	30	Other epidemic CC30	-	Sweden	Fitzgerald et al., 2001
MSA2754	30	Other epidemic CC30	-	Canada	Fitzgerald et al., 2001
MSA3095	30	Other epidemic CC30	1987	Canada	Fitzgerald et al., 2001
MSA537	30	Other epidemic CC30	1985	USA	Fitzgerald et al., 2001
PS10	30	Other epidemic CC30	-	USA	This study
PS11	30	Other epidemic CC30	-	USA	This study
PS13	30	Other epidemic CC30	-	USA	This study
PS14	30	Other epidemic CC30	-	USA	This study
PS17	30	Other epidemic CC30	-	USA	This study
PS2	30	Other epidemic CC30	-	USA	This study
PS20	30	Other epidemic CC30	-	USA	This study
PS4	30	Other epidemic CC30	-	USA	This study
PS5	30	Other epidemic CC30	-	USA	This study
PS6	30	Other epidemic CC30	-	USA	This study
PS7	30	Other epidemic CC30	-	USA	This study
PS8	30	Other epidemic CC30	-	USA	This study

Isolate	ST	Lineage	Isolation Year	Country	Reference
PS9	30	Other epidemic CC30	-	USA	This study
WW2703_97	30	Other epidemic CC30	1997	Germany	Robinson et al., 2005
WW2707_97	30	Other epidemic CC30	1997	Germany	Robinson et al., 2005
MSA1836	442	Other epidemic CC30	1972	USA	Fitzgerald et al., 2001
02_7810_P	novel ST30 slv	Other epidemic CC30	2002	Scotland	This study
03_7601_Q	novel ST30 slv	Other epidemic CC30	2003	Scotland	This study
03_9652_M	novel ST30 slv	Other epidemic CC30	2003	Scotland	This study
05_8345_P	novel ST30 slv	Other epidemic CC30	2005	Scotland	This study
09_1004_N	novel ST30 slv	Other epidemic CC30	2009	Scotland	This study
PS18	novel ST30 slv	Other epidemic CC30	-	USA	This study
08_8403_F	30	Southwest Pacific	2008	Scotland	This study
09_6292_C	30	Southwest Pacific	2009	Scotland	This study
11_1491_D	30	Southwest Pacific	2011	Scotland	This study
11_2558_B	30	Southwest Pacific	2011	Scotland	This study
11_3775_X	30	Southwest Pacific	2011	Scotland	This study
11_3902_X	30	Southwest Pacific	2011	Scotland	This study
11_4527_D	30	Southwest Pacific	2011	Scotland	This study
11_5404_G	30	Southwest Pacific	2011	Scotland	This study

Isolate	ST	Lineage	Isolation Year	Country	Reference
11_5558_L	30	Southwest Pacific	2011	Scotland	This study
11_5717_J	30	Southwest Pacific	2011	Scotland	This study
1973_07.17	30	Southwest Pacific	2007	Scotland	This study
7997_01	30	Southwest Pacific	2001	Scotland	Robinson et al., 2005
TCH_00.58	30	Southwest Pacific	2000	USA	Human Microbiome Project
WBG_99	30	Southwest Pacific	1999	Australia	Robinson et al., 2005
02_6340_B	39	ST39	2002	Scotland	This study
02_7657_Q	39	ST39	2002	Scotland	This study
MSA1827	39	ST39	1964	USA	Fitzgerald et al., 2001
MSA2335	39	ST39	-	Sweden	Fitzgerald et al., 2001
PS16	39	ST39	-	USA	This study

CC, clonal complex; ST, sequence type

Appendix 2


55.2798_55
55.2053_55
55.3224_55
e1410_62
M1016_62
M899_61
M876_61
M1015_62
58.362_58
65.20_65
68.397_68
66.1888_66
65.1322_65
58.424_58
68.81_68
69.172_69
69.412_69
1973_07.17
TCH_00.58
7997_01
VBG_99
sa17_06.50
1791_03.08
sa36_06.25
sa13_07.33
2449_07.17
6659_07.67
sa22_07.42
sa23_07.50
6636_07.67
sa35_07.50
2496_07.17
sa32_07.42
1712_07.08
sa12_07.67
sa14_06.50
sa9_07.67
sa10_07.75
sa11_07.42
sa37_07.33
sa29_99.17
sa30_07.75
sa16_06.75
sa3_97.92
MRSA252_97
1696_07.08
sa2_97.00
sa31_02.00
sa21_03.00
sa15_08.25
1227_07.08
9570_06.92
7206_07.75
5806_98.83
2589_07.17
sa38_05.83
3841_07.33
sa25_99.67
sa27_00.33
sa7_02.00
DEN4415_01
sa1_98.92
sa33_98.92
sa20_07.83
sa8_02.00
2880_07.25
sa34_06.83
sa19_06.00
sa18_07.50
sa28_02.42
sa24_07.50
sa6_93.00
HAR24_93
2483_97.75
sa4_97.67
75916_95
sa5_95.00
3366_99
sa26_00.25
7G_02
HU275_02
ED84_84.83
ED86_86.08
ED83_83.92
MN8_80
VW2703_97
A017934_97
0.61
0.92
0.46
0.98
0.36
0.98
0.15
0.16
0.16
0.37
0.99
0.1
0.22
0.22
0.98
0.34
20.0

Bayesian phylogeny of original CC30 dataset with nodes labelled with posterior support, and tips labelled with isolate names. Scale bar represents unit time in years.

Appendix 3