



*Citation for published version:*

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*', *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 23, pp. 9107-9112.  
<https://doi.org/10.1073/pnas.1202869109>

*DOI:*

[10.1073/pnas.1202869109](https://doi.org/10.1073/pnas.1202869109)

*Publication date:*

2012

*Document Version*

Peer reviewed version

[Link to publication](#)

## University of Bath

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# **Molecular Tracing of the Emergence, Adaptation and Transmission of Hospital-Associated MRSA**

Paul R. McAdam<sup>a</sup>, Kate E. Templeton<sup>b</sup>, Giles F. Edwards<sup>c</sup>, Matthew T. G. Holden<sup>d</sup>, Edward J. Feil<sup>e</sup>, David Aanensen<sup>f</sup>, Mark C. Enright<sup>g</sup>, Anne Holmes<sup>a</sup>, E. Kirsty Girvan<sup>c</sup>, Paul A. Godfrey<sup>h</sup>, Michael Feldgarden<sup>h</sup>, Angela M. Kearns<sup>i</sup>, Andrew Rambaut<sup>j,k</sup>, D. Ashley Robinson<sup>l</sup>, and J. Ross Fitzgerald<sup>a,1</sup>

<sup>a</sup>The Roslin Institute and Edinburgh Infectious Diseases, Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian, EH259RG, United Kingdom; <sup>b</sup>Microbiology, Royal Infirmary of Edinburgh, Edinburgh, United Kingdom; <sup>c</sup>Scottish MRSA Reference Laboratory, NHS Greater Glasgow and Clyde, Stobhill Hospital, Glasgow, United Kingdom; <sup>d</sup>The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom; <sup>e</sup>Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath, United Kingdom; <sup>f</sup>Department of Infectious Disease Epidemiology, Imperial College London, Norfolk Place, London, United Kingdom; <sup>g</sup>AmpliPhi Biosciences Corp., Colworth Science Park, Sharnbrook, Bedfordshire, United Kingdom; <sup>h</sup>Broad Institute of Harvard and MIT, Cambridge, Massachusetts, United States of America; <sup>i</sup>Microbiology Services Colindale, Health Protection Agency, London, United Kingdom; <sup>j</sup>Institute of Evolutionary Biology, University of Edinburgh, Ashworth Laboratories, Midlothian, United Kingdom; <sup>k</sup>Division of International Epidemiology and Population Studies, Fogarty International Center, National Institutes of Health, Bethesda, Maryland, United States of America; <sup>l</sup>Department of Microbiology, University of Mississippi Medical Center, Jackson, Mississippi, United States of America.

<sup>1</sup>To whom correspondence may be addressed; Email; [Ross.Fitzgerald@ed.ac.uk](mailto:Ross.Fitzgerald@ed.ac.uk)

JRF, PRM, KET GFE DAR designed research; PRM, MTGH, EJF, DA, MCE, AH, PAG, MF, AR, DAR performed research and analyzed data; GFE, EKG, AMK contributed new reagents or analytic tools; PRM and JRF wrote the paper.

Running title: Evolution of hospital-associated MRSA

Keywords: evolution, MRSA, hospital, phylogeography

Classification: BIOLOGICAL SCIENCES Microbiology

Manuscript: 19 pages text, 2 Figures.

Supporting Information: pages text, Figure, Tables

## Abstract

Hospital-associated infections caused by methicillin-resistant *Staphylococcus aureus* (MRSA) are a global health burden dominated by a small number of bacterial clones. The pandemic EMRSA-16 clone (ST36-II) has been widespread in UK hospitals for 20 years, but its evolutionary origin and the molecular basis for its hospital-association are unclear. We carried out a Bayesian phylogenetic reconstruction based on the genome sequences of 86 *S. aureus* isolates including 60 EMRSA-16 and 26 additional clonal complex 30 (CC30) isolates, collected from patients in 3 continents over a 53 year period. The 3 major pandemic clones to originate from the CC30 lineage, including phage type 80/81, South-West Pacific, and EMRSA-16, shared a most recent common ancestor that existed over 100 years ago while the hospital-associated EMRSA-16 clone is estimated to have emerged about 35 years ago. Our CC30 genome-wide analysis revealed striking molecular correlates of hospital- or community-associated pandemics represented by non-synonymous mutations and mobile genetic elements affecting antibiotic resistance and virulence. Importantly, phylogeographic analysis indicates that EMRSA-16 spread within the UK by transmission from large hospitals in London and Glasgow to regional health-care centers, implicating patient referrals as an important cause of nationwide transmission. Taken together, the high-resolution phylogenomic approach employed resulted in a new understanding of the emergence and transmission of a major MRSA clone and provided molecular correlates of its hospital adaptation. Similar approaches for hospital-associated clones of other bacterial pathogens may inform appropriate measures for controlling their intra- and inter-hospital spread.

## 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 immuno-compromised patients in hospitals, and healthy humans in the community.<sup>1</sup> Hospital-associated methicillin-resistant *S. aureus* (HA-MRSA) is represented by a small number of clones that rarely cause disease outside of the health-care setting, and which are characterized by resistance to beta-lactam antibiotics in addition to other front-line antimicrobials.<sup>2</sup> During the last 60 years, the *S. aureus* clonal complex 30 (CC30) has had a profound impact on global human health by giving rise to 3 pandemic waves and the toxic shock syndrome epidemic.<sup>3,4</sup> Furthermore, one study indicated that *S. aureus* isolates from life-threatening endocarditis infections are more likely to belong to CC30 than to other *S. aureus* lineages.<sup>5</sup> The first CC30 pandemic was caused by the 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 Panton-Valentine leukocidin (PVL) toxin.<sup>6-9</sup> The South-West Pacific clone (SWP) is a contemporary PVL-positive community-associated MRSA clone that has spread to several continents and which largely causes skin and soft tissue infections of otherwise healthy individuals.<sup>10,11</sup> 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.<sup>12</sup> 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, South-East Asia, South Africa, Australia, and North America.<sup>3,13-17</sup> 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.<sup>18,19</sup> Of note, EMRSA-16 prevalence has declined more rapidly than that of EMRSA-15, implying the existence

of unknown strain-dependent factors which confer increased susceptibility to hospital infection prevention and control measures.<sup>18,19</sup>

Despite its clinical importance, the evolution of the EMRSA-16 clone, in addition to the molecular basis for its success are poorly understood. Here we employ a phylogenomic approach to examine the diversity of the EMRSA-16 clone relative to other CC30 pandemic clones. The results provide an unprecedented level of resolution into the emergence and transmission of a major clone of MRSA revealing molecular correlates for its hospital association.

## Methods

**Bacterial isolates.** For genome sequencing, a total of 83 *S. aureus* isolates were selected to represent the breadth of genotypic diversity within the CC30 lineage sampled through time and space, with an emphasis on EMRSA-16 isolates in the UK (Table S1). The majority of isolates were typed by multi locus sequence typing (MLST) as ST30 (n=26) or its single locus variant (slv) ST36 (n=58), with one isolate identified as ST500, an slv of both ST30 and ST36. Genomic DNA was isolated from *S. aureus* as previously described.<sup>36</sup>

**Whole genome sequencing, mapping and alignment.** Whole-genome sequencing was carried out with the Illumina Genome Analyzer II, or the Roche 454 GS FLX platform. Adaptor sequences were trimmed from Illumina reads using the ea-utils FASTQ processing tool,<sup>20</sup> and low-quality reads were filtered out using the FASTX-toolkit.<sup>21</sup> Filtered Illumina reads were mapped to the ST36 MRSA252 genome sequence (accession number NC\_002952) using the BWA short read aligner with the Smith-Waterman alignment of unmapped mates disabled for paired end reads.<sup>22</sup> 454 reads were trimmed using the Biopython SeqIO module and mapped to MRSA252 using the BWA long read aligner.<sup>23</sup> 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, Acc. no. CM000952.1) were aligned using the progressiveMauve algorithm and gap positions removed.<sup>24</sup>

**Bayesian evolutionary analysis.** Bayesian analysis of evolutionary rates and divergence times was performed using BEAST v1.6.1<sup>25</sup> under the HKY model of nucleotide substitution. The SWP and Phage 80/81 clades were constrained together based on robust phylogenies determined using maximum-likelihood (Fig. SX), neighbour-joining and parsimony analyses. All isolates were dated based on year (and where known, month) of isolation. Markov Chain Monte Carlo (MCMC)

samples from 3 independent analyses each run for  $1.5 \times 10^8$  iterations, sampled every 1000 generations and the first 10% discarded as burnin, were combined for estimation of posterior probabilities. The relaxed lognormal molecular clock model was used, with a constant coalescent prior.<sup>26</sup> For an alignment of the UK isolates of the EMRSA-16 clone, phylogeographic distribution was examined using the discrete diffusion model with distance-informed priors.<sup>27</sup> Using city of isolation to construct the matrix of geographic locations resulted in an over-parameterized model. Therefore isolates were grouped by geographic region (London, South East, South and Central England, North, East and West Scotland), and mean distance between cities was calculated.



## Results and Discussion

**Phylogenetic and dating analysis of CC30 pandemics.** The core genome of the 86 CC30 isolates examined consisted of 2,381,276 bases (82% of the MRSA252 reference genome), containing a total of 4499 high confidence single nucleotide polymorphisms (SNP). We applied a Bayesian coalescent method using the relaxed lognormal molecular clock model to infer the phylogeny and the rate of molecular evolution of the CC30 lineage and its major clades. The phylogeny indicates 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 a paraphyletic clade represented by other CC30 epidemic (OCE) isolates with strong posterior support for the majority of nodes in the tree (Fig. 1). Importantly, parsimony analysis indicates a very low frequency of homoplasies across the phylogenetic tree (Consistency Index of 0.92) implying that it is an accurate depiction of the evolutionary relatedness of the CC30 strains examined. The mean nucleotide substitution rate within CC30 was  $1.42 \times 10^{-6}$  substitutions per site per year (95% HPDs  $1.05 \times 10^{-6}$ - $1.78 \times 10^{-6}$ ), and varied negligibly depending on the clock model, choice of tree prior and within each pandemic clade. Given the rate of molecular evolution determined for the CC30 lineage, we calculated the time of the most recent common ancestor (MRCA) for each of the three major clades that correspond to the 3 pandemic CC30 clones. The date for the MRCA of the 80/81 clone was estimated as 1937 (95% HPDs 1927-1945), the date for the MRCA of the SWP clone was estimated as 1968 (95% HPDs 1953-1984), and the date for the MRCA of the EMRSA-16 clone was estimated as 1975 (95% HPDs 1966-1983). The latter date precedes the first reports of EMRSA-16 identification in UK hospitals by about 18 years. Finally, the date for the MRCA of the entire CC30 lineage was estimated as 1845 (95% HPDs 1771-1908).

Previously, Robinson et al. employed 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.<sup>3</sup> However, consistent with the recent findings of DeLeo et al., our phylogenetic

analysis clearly demonstrates the independent origins of each of the three major CC30 pandemic clones that are estimated to have shared a MRCA which existed at least 100 years ago. Temporal analysis of bacterial epidemics may implicate contemporaneous environmental factors or human practises that promoted their emergence and expansion. For example, the phage type 80/81 clone is predicted to have emerged during a time of intensive penicillin usage which may have provided a selection for clonal expansion after acquisition of a  $\beta$ -lactamase plasmid.<sup>3,28,29</sup> 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*.<sup>30-32</sup> Early strains of the emergent clones may have existed for some time prior to acquiring mutations or mobile genetic elements (MGE) associated with clonal expansion.

**Multiple independent acquisitions of *pvl* occurred during the evolution of CC30 pandemic clones.** 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 which may have contributed to clone emergence and pathogenesis.<sup>31,33-35</sup> PVL has been a marker for community-associated clones of *S. aureus* associated with skin and soft tissue infections or severe necrotizing pneumonia<sup>36</sup>, though increased identification of PVL-positive strains associated with nosocomial infections weakens this correlation.<sup>37,38</sup> We examined the genomes of the CC30 isolates included in the current study for the presence of the *pvl* locus (*lukS-PV* and *lukF-PV*) and found that 19 of 21 isolates belonging to the community-associated phage type 80/81 and SWP clones (Fig. 1). The PVL toxin is encoded by temperate phage which can be differentiated into distinct morphological groups, based on their elongated- or icosahedral-head types.<sup>39</sup> Sequence analysis of the small and large terminase subunits of the PVL phage revealed that 3 of 4 SWP clone isolates contained the elongated-head phage type, while the remaining SWP strain, and the phage type 80/81 clone each had the icosahedral-head phage type. Phylogenetic analysis of

the small and large terminase subunit genes indicates the close relatedness of the icosahedral phage heads of the SWP and phage type 80/81 clones (Fig S3) 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. Subsequently, the phage was replaced by the elongated-head phage type in some isolates of the SWP clone (Fig. S3). Overall, these data suggest that PVL has a long residency with some community-associated *S. aureus* strains having been maintained in some CC30 clades since a likely acquisition event which occurred over 100 years ago. These data are consistent with a central role for PVL in the success of some community-associated *S. aureus* strains.<sup>40</sup>

### **A single acquisition of *tst* led to CC30 strains responsible for the toxic shock syndrome**

**epidemic.** Previous studies have demonstrated that the majority of cases of menstrual toxic shock syndrome (TSS) are caused by a single clone that corresponds to CC30.<sup>41,42</sup> However, the distribution of the *tst* gene encoding the toxic shock syndrome toxin-1 (TSST-1) among CC30 subclades has not been previously examined. We determined that the *tst* gene is harbored by the staphylococcal pathogenicity island 2 (SaPI2) and is present in 55 of 66 (83%) of the isolates from the clade represented by EMRSA-16 and the other epidemic CC30 clade (Fig. 1), but is not present in any of 21 (0%) of the isolates from the clades represented by the SWP and 80/81 clones. These data indicate 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 (3 or fewer SNPs in 14.7 kb SaPI2) strongly suggests that a single acquisition event occurred in an ancestor which existed at least 50 years ago, prior to the differentiation of EMRSA-16 from the other contemporary CC30 clones (Fig. 1) and 10 to 140 years before the TSS epidemic of the 1970/80s. These dating estimates are consistent with previous studies which provided evidence that the TSS epidemic was caused by already widely disseminated TSST-1-positive strains rather than rapid clonal expansion of a single strain which had acquired a fitness

mutation (eg SaPI2 acquisition).<sup>41</sup>

**Identification of mutations which correlate with the hospital habitat of EMRSA-16.** The hospital setting provides a unique array of insults which target bacteria including numerous classes of antibiotics used in treatment, and routine use of disinfectants, and detergents. However, in contrast to the community setting, the hospital provides a continuous supply of immunocompromised human hosts, which offer plentiful opportunities for infection and transmission. Our CC30 genome-wide analysis indicates that 58 of 60 EMRSA-16 isolates contained the *SCCmecII* element (Fig. 1). Acquisition of *SCCmecII* was a critical genetic event in the evolution of the EMRSA-16 clone as a hospital-associated antibiotic-resistant clone refractory to treatment with  $\beta$ -lactam antibiotics. Of note, the type II *SCCmec* element has been demonstrated to reduce the toxicity of MRSA CC30 strains in comparison to methicillin-sensitive CC30 strains, by preventing normal stationary phase induction of the *agr* system, leading to decreased expression of cytolytic toxins.<sup>43,44</sup> It is speculated that this reduction in energy requirement could compensate for the metabolically-costly maintenance of a large *SCCmec* element and its associated methicillin resistance, but which would be likely to lead to reduced fitness outside of the hospital setting.<sup>45</sup>

In addition to *SCCmecII*, we identified a number of non-synonymous mutations specific to EMRSA-16 on loci previously demonstrated to influence antibiotic resistance, which are likely to have been the result of selective pressures prevalent in hospitals. 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.<sup>46-49</sup> Of note, a selection of 16 EMRSA-16 isolates tested were all fluoroquinolone resistant (data not shown). In addition, non-synonymous mutations were identified in additional 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 for the multi-drug efflux transporter NorA in 50 of 60 EMRSA-16 isolates.

Finally plasmids encoding resistance to quaternary ammonium compounds were found in 7 of 60 EMRSA-16 isolates (FigS2).<sup>12</sup>

Recently, DeLeo and colleagues carried out a comparative genomic analysis of 9 CC30 isolates including a single EMRSA-16 isolate (MRSA252). The authors identified mutations in *agr* and *hla* genes which occurred in the ancestor of contemporary CC30 isolates which caused reduced virulence in mouse models of infection, leading them to infer that the mutations have contributed to the hospital-association of contemporary CC30 isolates.<sup>4</sup> However, isolates which belong to the other CC30 epidemic clone, defined in the current study as a paraphyletic clade (Fig. 1), which have been isolated from episodes of severe community-associated infections of healthy humans including the TSS epidemic, also contain the same *agr* and *hla* mutations (Fig. 1 and Table S2).<sup>50</sup> We therefore suggest that while *agr* and *hla* mutations may indeed influence the capacity of CC30 isolates to cause certain types of infection, they are likely not sufficient to explain the hospital restriction of the EMRSA-16 clone since its differentiation from the other CC30 epidemic clone.

We identified a number of mutations which occurred on the branch leading to the EMRSA-16 clone (Table SX). It should be pointed out that some of the mutations are likely to be the result of fixation due to genetic drift prior to clonal expansion rather than the result of selective pressures encountered in hospitals (Fig SX). However, several of the identified mutations are predicted to have functional consequences which could impact on virulence. For example, 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.<sup>50</sup> This is consistent with the non-pigmented colony phenotype of EMRSA-16 strains noted in the current study (data not shown), and described in the earliest reports of the EMRSA-16 clone.<sup>51</sup> Bacterial carotenoids confer resistance to oxidative killing during phagocytosis, and staphyloxanthin-deficient *S. aureus* mutants have diminished virulence in animal models of

infection.<sup>50</sup> The loss of metabolically-expensive pigment production may in part compensate for the energy cost of maintaining the large *SCCmecII* element mediating antibiotic resistance.<sup>31,52</sup> 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.

In addition non-synonymous mutations were identified in genes encoding proteins involved in virulence such as the virulence gene regulator CcpA, and (for a proportion of EMRSA-16 strains) cell wall-associated proteins SasH, IsdB, Fib, and Ebh. While the examples discussed are selected based on known or implicated roles in virulence, we cannot dismiss 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 (Table SX).

**EMRSA-16 has spread within the UK by transmission from hospitals in major cities to regional centers.** 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.<sup>25</sup> Bayesian phylogenetic analysis of a subset of the genome sequence data which included the UK EMRSA-16 isolates only (58 of 60 EMRSA-16 isolates) resolved several sub-clades consisting of isolates from proximal UK geographic locations, consistent with the existence of EMRSA-16 strains which are endemic to particular hospitals or regions (Fig. 2). 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 (Fig. 2). 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 (Fig. 2). However, isolates from London and Glasgow are widely distributed among clusters of closely-related isolates from regional hospitals, consistent with hospitals in major population centers acting as reservoir for EMRSA-16 UK transmissions. To examine this observation further we employed

the discrete phylogeographic diffusion model implemented in BEAST, with isolates grouped by geographic region (London, South East, South and Central England, North, East and West Scotland), and mean distance between cities calculated. An alternative approach employing a matrix of geographic locations based on city of isolation resulted in an over-parameterized model (Fig 2)<sup>53</sup>. Using this approach, we identified statistical support for London as a source of EMRSA-16 transmission events to South and South-East regions of England (Bayes Factors 3.00 and 4.36 respectively). In addition, Glasgow was identified as a reservoir for transmission of EMRSA-16 to surrounding population centers in the North and East of Scotland (Bayes Factors 4.04 and 7.16 respectively). The dataset is limited by the number of hospitals sampled in different regions. For example, it would be interesting to examine the dynamics between regional networks within the city of London. Nonetheless, these data 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 which used a simulation model to estimate high transmission rates between large hospitals and long term care facilities (REF). 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 centers have been contributing factors to the spread of MRSA throughout the UK.<sup>53</sup> These data could inform the design of infection control protocols, such as decolonization of patients prior to transfer from large hospitals, in order to limit inter-hospital transmission as a major driving force for epidemics.

**Concluding comments.** The capacity to rapidly sequence the genome of large numbers of bacterial isolates is revolutionizing the study of bacterial populations leading to unparalleled insights into bacterial epidemics. By employing a high-resolution phylogenomic approach, we have provided broad new insights into the emergence and transmission of a major hospital-associated MRSA clone which may be used to inform control methods. In addition, we have identified genetic events which correlate with its adaptation to the hospital environment, 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 of epidemics, and thereby inform the rational design of methods for controlling their inter- and intra-hospital spread.



## **Acknowledgements**

The authors are grateful to R. Goering, H. de Lencastre and A. Shore for providing *S. aureus* isolates, and to ARK-Genomics, Roslin Institute for sequencing services. The project was supported by a grant from the Chief Scientist's Office Scotland, a Doctoral Training Grant from the Medical Research Council (UK), the Biotechnology Biological Sciences Research Council (UK), contract number HHSN27220090018C from the National Institute of Allergy and Infectious Diseases (NIH, USA), grant number GM080602 from the National Institute of General Medical Sciences (NIH, USA), and The Wellcome Trust grant 098051.

## References

- 1 Wertheim HFL, Melles DC, Vos MC, *et al.* The role of nasal carriage in *Staphylococcus aureus* infections. *The Lancet infectious diseases* 2005; **5**: 751–62.
- 2 Chambers HF, DeLeo FR. Waves of resistance: *Staphylococcus aureus* in the antibiotic era. *Nature reviews Microbiology* 2009; **7**: 629–41.
- 3 Robinson DA, Kearns AM, Holmes A, *et al.* Re-emergence of early pandemic *Staphylococcus aureus* as a community-acquired methicillin-resistant clone. *Lancet* 2005; **365**: 1256–8.
- 4 Altemeier WA, Lewis SA, Schlievert PM, *et al.* *Staphylococcus aureus* associated with toxic shock syndrome: phage typing and toxin capability testing. *Annals of internal medicine* 1982; **96**: 978–82.
- 5 Nienaber JJC, Sharma Kuinkel BK, Clarke-Pearson M, *et al.* Methicillin-Susceptible *Staphylococcus aureus* Endocarditis Isolates Are Associated With Clonal Complex 30 Genotype and a Distinct Repertoire of Enterotoxins and Adhesins. *The Journal of infectious diseases* 2011; **204**: 704–13.
- 6 Blair JE, Carr M. Distribution of Phage Groups of *Staphylococcus aureus* in the Years 1927 through 1947. *Science (New York, NY)* 1960; **132**: 1247–8.
- 7 Nahmias A, Sakurai N, Blumberg R, Goe Ge A, Sulzer C. The *Staphylococcus* 80/81 complex: epidemiological and laboratory observations. *The Journal of infectious diseases* 1961; **109**: 211–22.
- 8 Donahue JA, Baldwin JN. Hemolysin and leukocidin production by 80/81 strains of *Staphylococcus aureus*. *The Journal of infectious diseases* 1966; **116**: 324–8.
- 9 Goldie DJ, Alder VG, Gillespie WA. Changes in the drug resistance of *Staphylococcus aureus* in a non-hospital population during a 20-year period. *Journal of clinical pathology* 1971; **24**: 44–7.
- 10 Collignon P, Gosbell I, Vickery A, Nimmo G, Stylianopoulos T, Gottlieb T. Community-acquired methicillin-resistant *Staphylococcus aureus* in Australia. Australian Group on Antimicrobial Resistance. *Lancet* 1998; **352**: 145–6.
- 11 Alesana-Slater J, Ritchie SR, Heffernan H, *et al.* Methicillin-resistant *Staphylococcus aureus*, Samoa, 2007–2008. *Emerging infectious diseases* 2011; **17**: 1023–9.
- 12 DeLeo FR, Kennedy AD, Chen L, *et al.* Molecular differentiation of historic phage-type 80/81 and contemporary epidemic *Staphylococcus aureus*. *Proceedings of the National Academy of Sciences* 2011; : 1111084108-.
- 13 Ho P-L, Chow K-H, Lo P-Y, Lee K-F, Lai EL. Changes in the epidemiology of methicillin-resistant *Staphylococcus aureus* associated with spread of the ST45 lineage in Hong Kong. *Diagnostic microbiology and infectious disease* 2009; **64**: 131–7.
- 14 Jansen van Rensburg MJ, Eliya Madikane V, Whitelaw A, Chachage M, Haffejee S, Gay Elisha B. The dominant methicillin-resistant *Staphylococcus aureus* clone from hospitals in Cape Town has an unusual genotype: ST612. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* 2011; **17**: 785–92.
- 15 Carleton HA, Diep BA, Charlebois ED, Sensabaugh GF, Perdreau-Remington F. Community-adapted methicillin-resistant *Staphylococcus aureus* (MRSA): population dynamics of an expanding community reservoir of MRSA. *The Journal of infectious diseases* 2004; **190**: 1730–8.
- 16 Scicluna EA, Shore AC, Thürmer A, *et al.* Characterisation of MRSA from Malta and the description of a Maltese epidemic MRSA strain. *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology* 2010; **29**: 163–70.

- 17 Pérez-Roth E, Lorenzo-Díaz F, Batista N, Moreno A, Méndez-Alvarez S. Tracking methicillin-resistant *Staphylococcus aureus* clones during a 5-year period (1998 to 2002) in a Spanish hospital. *Journal of clinical microbiology* 2004; **42**: 4649–56.
- 18 Wyllie DH, Walker AS, Miller R, *et al.* Decline of methicillin-resistant *Staphylococcus aureus* in Oxfordshire hospitals is strain-specific and preceded infection-control intensification. *BMJ open* 2011; **1**: e000160.
- 19 Ellington MJ, Hope R, Livermore DM, *et al.* Decline of EMRSA-16 amongst methicillin-resistant *Staphylococcus aureus* causing bacteraemias in the UK between 2001 and 2007. *Journal of Antimicrobial Chemotherapy* 2010; **65**: 446–8.
- 20 Aronesty E. ea-utils. 2011; : <http://code.google.com/p/ea-utils/>.
- 21 Gordon A. FASTX-toolkit. 2010; : [http://hannonlab.cshl.edu/fastx\\_toolkit/index.html](http://hannonlab.cshl.edu/fastx_toolkit/index.html).
- 22 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* 2009; **25**: 1754–60.
- 23 Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* 2010; **26**: 589–95.
- 24 Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS one* 2010; **5**: e11147.
- 25 Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology* 2007; **7**: 214.
- 26 Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS biology* 2006; **4**: e88.
- 27 Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. *PLoS computational biology* 2009; **5**: e1000520.
- 28 Gould JC. Environmental penicillin and penicillin-resistant *Staphylococcus aureus*. *The Lancet* 1958; **271**: 489–93.
- 29 Rountree PM, Asheshov EH. Further Observations on Changes in the Phage-Typing Pattern of Phage Type 80/81 *Staphylococci*. *Microbiology* 1961; **26**: 111–22.
- 30 Smyth DS, McDougal LK, Gran FW, *et al.* Population structure of a hybrid clonal group of methicillin-resistant *Staphylococcus aureus*, ST239-MRSA-III. *PLoS one* 2010; **5**: e8582.
- 31 Harris SR, Feil EJ, Holden MTG, *et al.* Evolution of MRSA during hospital transmission and intercontinental spread. *Science* 2010; **327**: 469–74.
- 32 Gray RR, Tatem AJ, Johnson JA, *et al.* Testing spatiotemporal hypothesis of bacterial evolution using methicillin-resistant *Staphylococcus aureus* ST239 genome-wide data within a Bayesian framework. *Molecular biology and evolution* 2011; **28**: 1593–603.
- 33 Mutreja A, Kim DW, Thomson NR, *et al.* Evidence for several waves of global transmission in the seventh cholera pandemic. *Nature* 2011. doi:10.1038/nature10392.
- 34 Croucher NJ, Harris SR, Fraser C, *et al.* Rapid pneumococcal evolution in response to clinical interventions. *Science (New York, NY)* 2011; **331**: 430–4.
- 35 Hendriksen RS, Price LB, Schupp JM, *et al.* Population Genetics of *Vibrio cholerae* from Nepal in 2010: Evidence on the Origin of the Haitian Outbreak. *mBio* 2011; **2**: e00157–11-e00157–11.

- 36 Diep BA, Chan L, Tattevin P, *et al.* Polymorphonuclear leukocytes mediate *Staphylococcus aureus* Panton-Valentine leukocidin-induced lung inflammation and injury. *Proceedings of the National Academy of Sciences of the United States of America* 2010; **107**: 5587–92.
- 37 Rossney AS, Shore AC, Morgan PM, Fitzgibbon MM, O’Connell B, Coleman DC. 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. *Journal of clinical microbiology* 2007; **45**: 2554–63.
- 38 Popovich KJ, Weinstein RA, Hota B. Are community-associated methicillin-resistant *Staphylococcus aureus* (MRSA) strains replacing traditional nosocomial MRSA strains? *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2008; **46**: 787–94.
- 39 Ma XX, Ito T, Kondo Y, *et al.* Two different Panton-Valentine leukocidin phage lineages predominate in Japan. *Journal of clinical microbiology* 2008; **46**: 3246–58.
- 40 Diep BA, Sensabaugh GF, Somboonna N, Somboona NS, Carleton HA, Perdreau-Remington F. Widespread skin and soft-tissue infections due to two methicillin-resistant *Staphylococcus aureus* strains harboring the genes for Panton-Valentine leukocidin. *Journal of clinical microbiology* 2004; **42**: 2080–4.
- 41 Fitzgerald JR, Sturdevant DE, Mackie SM, Gill SR, Musser JM. Evolutionary genomics of *Staphylococcus aureus*: insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. *Proceedings of the National Academy of Sciences of the United States of America* 2001; **98**: 8821–6.
- 42 Peacock SJ, Moore CE, Justice A, *et al.* Virulent combinations of adhesin and toxin genes in natural populations of *Staphylococcus aureus*. *Infection and immunity* 2002; **70**: 4987–96.
- 43 Rudkin JK, Edwards AM, Bowden MG, *et al.* Methicillin Resistance Reduces the Virulence of Healthcare-Associated Methicillin-Resistant *Staphylococcus aureus* by Interfering With the agr Quorum Sensing System. *The Journal of infectious diseases* 2012. doi:10.1093/infdis/jir845.
- 44 Collins J, Rudkin J, Recker M, Pozzi C, O’Gara JP, Massey RC. Offsetting virulence and antibiotic resistance costs by MRSA. *The ISME journal* 2010; **4**: 577–84.
- 45 Trong HN, Prunier A-L, Leclercq R. Hypermutable and fluoroquinolone-resistant clinical isolates of *Staphylococcus aureus*. *Antimicrobial agents and chemotherapy* 2005; **49**: 2098–101.
- 46 Łeski TA, Tomasz A. 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. *Journal of bacteriology* 2005; **187**: 1815–24.
- 47 Banerjee R, Gretes M, Harlem C, Basuino L, Chambers HF. A mecA-negative strain of methicillin-resistant *Staphylococcus aureus* with high-level  $\beta$ -lactam resistance contains mutations in three genes. *Antimicrobial agents and chemotherapy* 2010; **54**: 4900–2.
- 48 Hiron A, Falord M, Valle J, Débarbouillé M, Msadek T. 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. *Molecular microbiology* 2011; **81**: 602–22.
- 49 Couto I, Costa SS, Viveiros M, Martins M, Amaral L. Efflux-mediated response of *Staphylococcus aureus* exposed to ethidium bromide. *The Journal of antimicrobial chemotherapy* 2008; **62**: 504–13.
- 50 Liu GY, Essex A, Buchanan JT, *et al.* *Staphylococcus aureus* golden pigment impairs neutrophil killing and promotes virulence through its antioxidant activity. *The Journal of experimental medicine* 2005; **202**: 209–15.
- 51 Epidemic methicillin resistant *Staphylococcus aureus*. *Communicable disease report CDR weekly* 1997; **7**: 191.

- 52 Beres SB, Carroll RK, Shea PR, *et al.* Molecular complexity of successive bacterial epidemics deconvoluted by comparative pathogenomics. *Proceedings of the National Academy of Sciences of the United States of America* 2010; **107**: 4371–6.
- 53 Murchan S, Aucken HM, O'Neill GL, Ganner M, Cookson BD. Emergence, spread, and characterization of phage variants of epidemic methicillin-resistant *Staphylococcus aureus* 16 in England and Wales. *Journal of clinical microbiology* 2004; **42**: 5154–60.

[Figure legends]

**Fig. 1. The CC30 lineage is divided into multiple distinct clades characterized by the presence**

**of different toxin and antibiotic resistance determinants** Bayesian phylogenetic reconstruction

of the CC30 lineage using all sites in the core genome. Blue, green, red, and grey shading

correspond to the 80/81, South-West Pacific, EMRSA-16, and other epidemic CC30 clones

respectively. Presence of the *pvl* locus 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

methicillin-resistant isolates. Branch lengths are scaled according to time-scale bar. All nodes have

posterior probability support >80 unless labeled.

**Fig. 2. EMRSA-16 has been transmitted from hospitals in major population centers to**

**regional centers. A.** Bayesian phylogenetic reconstruction of UK EMRSA-16 isolates. Terminal

branches representing London and Glasgow isolates are colored in red and blue, respectively. Black

branches depict isolates from other locations. Branches are scaled with time (years). Gray shading

indicates examples of geographically-restricted subclades a and b. B Map of UK indicating the

sampled regions, Central England (CE), Southern England (SE) and South-East England (SEE),

North Scotland (NS), East Scotland (ES), and West Scotland (WS), with Glasgow (G) and London

(L) depicted in blue and red, respectively.



Fig. 2

