Whole genome sequencing reveals the contribution of long-term carriers in Staphylococcus aureus outbreak investigation

Article (Accepted Version)


This version is available from Sussex Research Online: http://sro.sussex.ac.uk/id/eprint/68044/

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:
Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.
Whole Genome Sequencing reveals the contribution of long-term carriers in Staphylococcus aureus outbreak investigation

Gordon NC1,2*, Pichon B3, Golubchik T1,2, Wilson DJ1,2, Paul J4, Blanc DS5, Cole K4, Collins J6, Cortes N7, Cubbon M8, Gould FK8, Jenks PJ9, Llewelyn M8,10, Nash JQ11, Orendi JM12, Paranatham K13, Price J4, Senn L5, Thomas HL13, Wyllie S7, Crook DW1,2,13,14, Peto TEA1,2,14, Walker AS1,2,14#, Kearns AM3#

1. National Institute for Health Research Oxford Biomedical Research Centre, John Radcliffe Hospital, Oxford, UK
2. Nuffield Department of Medicine, University of Oxford, UK
4. Public Health England, Royal Sussex County Hospital, Brighton, UK.
5. Lausanne University Hospital, Service of Preventative Medicine, Lausanne, Switzerland
6. Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle, UK
7. Portsmouth Hospitals NHS Trust, Portsmouth, UK
8. Brighton and Sussex University Hospitals NHS Trust, Brighton, UK
9. Plymouth Hospitals NHS Trust, Plymouth, UK
10. Brighton and Sussex Medical School, Falmer, UK
11. East Kent Hospitals NHS Foundation Trust, Canterbury, UK
12. Royal Stoke University Hospital, University Hospitals of North Midlands NHS Trust, Stoke-on-Trent, UK
14. The National Institute for Health Research Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance at University of Oxford

*Corresponding author; # contributed equally

*Dr NC Gordon
Nuffield Department of Medicine
John Radcliffe Hospital
Oxford OX3 9DU
Tel: +44(0)1865-220883 Fax: +44(0)1865-22195
E-mail: claire.gordon@ndm.ox.ac.uk
Abstract

Whole genome sequencing (WGS) makes it possible to determine the relatedness of bacterial isolates at high resolution, helping to characterise outbreaks. However, for *Staphylococcus aureus*, accumulation of within-host diversity during carriage might limit interpretation of sequencing data.

In this study, we hypothesised the converse: that within-host diversity can in fact be exploited to reveal the involvement of long-term carriers (LTCs) in outbreaks. We analysed WGS data from 20 historical outbreaks, and applied phylogenetic methods to assess genetic relatedness and estimate time to most recent common ancestor (TMRCA). Findings were compared with the routine investigation results and epidemiological evidence.

Outbreaks with epidemiological evidence for an LTC source had a mean estimated TMRCA (adjusted for outbreak duration) of 243 days (95% CI 143-343), compared with 55 days (28-81) for outbreaks lacking epidemiological evidence for an LTC (p=0.004). A threshold of 156 days predicted LTC involvement with a sensitivity of 0.875 and a specificity of 1.

We also found 6/20 outbreaks included isolates with differing antimicrobial susceptibility profiles, however, these had only modestly increased pairwise diversity (mean 17.5 single nucleotide variants (SNVs) (95% CI 17.3-17.8) vs 12.7 SNVs (12.5-12.8)) compared with isolates with identical antibiograms (p<0.0001). Additionally, for 2 outbreaks, WGS identified 1 or more isolates which were genetically distinct despite having the outbreak PFGE pulstype.

Duration-adjusted TMRCA allowed the involvement of LTCs in outbreaks to be identified and could be used to decide whether screening for long-term carriage (e.g. in healthcare workers) is warranted. Requiring identical antibiograms to trigger investigation could miss important contributors to outbreaks.
Introduction

To manage *Staphylococcus aureus* outbreaks effectively, infection control practitioners need to determine the relatedness of isolates from suspected cases. Whole genome sequencing (WGS) has shown superior resolution compared with standard typing techniques (*spa*, pulsed field gel electrophoresis (PFGE)) when used for individual outbreaks (1-4), and can also provide additional information about resistance, pathogenicity and population structure (5-8). However, it has been argued that the accumulation of within-host diversity during *S. aureus* carriage could result in erroneous inferences about transmission. This has been cited as a potential weakness in applying sequencing to *S. aureus* outbreaks, and may lead to misinterpretation of genuine transmission routes (1, 9, 10).

However, rather than within-host diversity being a limitation on sequencing-based outbreak investigation, it could in fact be exploited to determine whether a long-term carrier is implicated in maintaining an outbreak. This information could be used by infection control practitioners when considering whether or not to deploy extended screening (e.g. of healthcare workers).

In this study, we tested the hypothesis that WGS can be used to predict the presence of a long-term carrier as an outbreak source. First, we examined individuals with newly acquired *S. aureus* nasal carriage to ascertain whether diversity is present at acquisition or develops over time. Next, we analysed 20 *S. aureus* outbreaks, previously investigated using standard typing techniques, to assess the added utility of WGS. Finally, we compared WGS with epidemiological data to determine whether the presence of a long-term carrier maintaining the outbreak could be inferred from the WGS data.

Results

Comparison of within-host diversity in newly-acquired and long-term carriage

Eight subjects were identified with ≥3 consecutive bi-monthly negative nasal swabs, followed by ≥1 year of swabs consistently positive for *S. aureus*. All isolates were MSSA,
representing 7 spa-types, 5 sequence-types and 4 clonal complexes. Median time from first
to last positive sample was 490 days (range 358-727). In total, 135 isolates were
successfully sequenced from 16 samples. One isolate (case 1219, early sample) failed
quality checks and was excluded.

In 6/8 subjects, there was a significant increase in mean pairwise diversity (MPWD) between
the first and last samples (p<0.05; figure 1). In one participant (1236) the increase was not
significant (p=0.52), and for one (1375), there was a decrease which was marginally
significant (p=0.07). Overall, MPWD increased from 0.88 single nucleotide variants (SNVs)
(95%CI 0.65-1.11) to 3.30 (2.92-3.68) between first and last samples (p<0.001). Analysis of
the phylogenetic trees (see supplementary data) showed highly clonal early populations, and
in 2 participants only a single strain was observed. One individual (1219) had a more diverse
early sample (MPWD 4.57, 95%CI 3.10-6.04) compared with the other participants. This
subject’s first positive swab was at month 12, and they had completed a course of co-
amoxiclav one day before their final negative swab. It is therefore possible that this was a
false negative due to antibiotic suppression, meaning that there may have been up to four
months of carriage prior to the first positive swab, accounting for the increased diversity.

Two participants (1218 and 1219) shared the same address, and had isolates of the same
spa-type. Participant 1219 (donor) became positive two months before participant 1218
(recipient). On direct comparison of both early populations, we found that the recipient had
an entirely clonal initial population, identical to 4/8 of the donor’s strains (supplementary
data).

For an additional 13 participants positive at study entry, within-host diversity as measured by
MPWD ranged from 0 SNVs (3 individuals) to 26 SNVs. This may be due to differences in
acquisition time to time of first sample, which is unknown for these individuals.
Outbreak characteristics

Twenty outbreaks were included in the study (table 1). Fourteen (70%) were hospital-associated: 5 neonatal units, 4 general wards, 1 surgical unit, 2 maternity units, and 2 involved multiple wards or hospital sites. Six (30%) were community-associated: 4 households, 1 nursing home and 1 school. Reasons for instigating an outbreak investigation were: increase in MRSA carriage (8 outbreaks); Panton-Valentine Leukocidin (PVL)-producing skin/soft tissue infection (7 outbreaks); surgical site infections (3 outbreaks); MRSA bacteraemia (1 outbreak) and staphylococcal scalded skin syndrome (1 outbreak).

Three (15%) were due to MSSA, and 17 (85%) to MRSA. The median number of outbreak cases was 7 (IQR 5-9). Median duration was 72 days (IQR 44-188).

Overall, isolates from 391 cases were sequenced. Nine (2.3%) were from health care workers (HCWs), the remainder being from patients or household members. Outbreak samples represented 9 clonal complexes, 11 sequence-types and 12 spa-types.

Phylogenetic analysis of outbreaks

Phylogenetic trees for each outbreak are provided in the supplementary data. Two outbreaks had isolates which were equally or more distant than comparator isolates, despite having the outbreak pulsotype: outbreak D (one isolate 53 SNVs from index case compared with 21) and outbreak S (two isolates 49 and 46 SNVs from index case compared with 46). These were therefore considered to be sporadic, non-outbreak isolates, and were excluded from further analysis.

The overall MPWD across all outbreak sample pairs for the remaining 388 isolates was 13.8 SNVs (95%CI 13.6-13.9), compared with 4444 SNVs for non-outbreak spa-matched pairs (95%CI 2492-6395) and 30192 SNVs for non-outbreak isolates from the same units (95%CI 29781-30603). All outbreak isolates were ≤30 SNVs from the index case. 381/388 (98%) were ≤10 SNVs from their nearest neighbour. The 7 more distant isolates came from
outbreaks lasting more than 6 months (B, G and S). All isolates were mapped to a standard reference genome: mapping to an alternative reference strain (performed for 6 outbreaks) yielded only 2 additional SNVs overall (see supplementary data), with no effect on topology.

Time to most recent common ancestor (TMRCA) and long-term carriers

Twelve outbreaks (60%) had epidemiological evidence of a long-term carrier (LTC): 3 included cases with recurrent staphylococcal disease, in 5 an LTC was suspected due to non-overlapping ward stays, and in 4, at least one case had post-outbreak long-term carriage (figure 2). The pairwise distances between isolates from outbreaks with evidence for an LTC ranged from 0 to 46 SNVs, compared with 0 to 10 SNVs for outbreaks with no evidence for an LTC (table 2). Mean duration-adjusted TMRCA for outbreaks with a suspected or proven LTC was 243 days (95% CI 143-343) compared with 55 days (28-81) for outbreaks with no evidence for an LTC (p=0.004, figure 2). Excluding post-outbreak carriage, analysis of the receiver-operating-characteristic curve gave an AUC of 0.953 (95% CI 0.851-1). Using the Youden index to select the optimal threshold gave a cut-off value of 156 days, with a sensitivity of 0.875 and a specificity of 1.

Relationship between PFGE pulsotype / antibiogram and SNV distance

Five outbreaks contained isolates differing by ≥1 band from the index case on PFGE. MPWD between outbreak isolates with identical PFGE pulsotypes was 13.6 SNVs (13.4-13.7), compared with 17.3 (17.0-17.6) between isolates with differing pulsotypes (p<0.0001).

In 6/20 outbreaks, antimicrobial susceptibility differed across isolates, confirmed by the presence / absence of mobile resistance determinants identified using BLAST, however, these clearly belonged to the outbreak on phylogenetic analysis. MPWD between isolates sharing an antibiogram was 12.7 SNVs (95% CI 12.5-12.8) compared with 17.5 (17.3-17.8) for isolates with differing antibiograms (p<0.001), although a substantial number of isolate pairs with different antibiograms had 0 SNVs between their core genomes (figure 4).
For other factors potentially related to outbreak diversity, there was no evidence of association between MPWD and any of: outbreak duration, reason for investigation, epidemiological setting or MRSA phenotype (p>0.05).

**Discussion**
We have tested the use of WGS for *S. aureus* outbreak investigation using 20 outbreaks. By comparing observed outbreak SNV distances with non-outbreak and spa/MLST specific diversity, we were able to distinguish outbreak from non-outbreak strains.

Our observation of minimal diversity in recent acquisition of nasal carriage is reassuring for the application of WGS data to outbreaks. For the donor-recipient pair, we observed a narrow transmission bottleneck, with a clonal founding population despite a diverse donor population. Although this is a single case, the findings are supported by the minimal diversity seen in the early samples for the majority of carriage study subjects, and further evidence for a narrow transmission bottleneck is provided by the relatively short SNV distances observed across the outbreaks. Taken together, these findings suggest that, in an acute short term outbreak, there will be insufficient time for diversity to accumulate.

If WGS is to be used routinely for outbreak investigation, these findings provide evidence that single colony sequencing is likely to identify clusters reliably in this context, allowing ease of interpretation and ensuring that WGS remains an affordable alternative to standard typing, as a requirement for sequencing multiple colonies per case, as implied by previous investigators (1, 10), would rapidly escalate costs and render WGS too expensive for routine use.

Previous carriage studies have found greater distances than seen here (9, 11), however, these did not account for estimated time of acquisition. We postulate that the existence of a
significant cloud of diversity (4, 12) may be a marker of long-term carriage, and therefore, in outbreaks, higher diversity may indicate the involvement of an LTC, with outbreak diversity reflecting the donor cloud.

In support of this, we observed a significant difference in duration-adjusted TMRCA between outbreaks with and without evidence of an LTC. The longest TMRCAs were in hospital outbreaks with indirect links between cases (i.e. non-overlapping ward stays). The likelihood of "missed" cases in these outbreaks was considered low due to enhanced screening, and the most likely reason for the reoccurrence of the outbreak strain was thought by the investigating teams to be either re-introduction from the community (outbreak G) or a staff member with long-term carriage (outbreaks A, I, N and S). Staff carriage was proven in one outbreak (by sampling and subsequent termination of the outbreak on their exclusion), but in the remaining outbreaks HCWs were either not sampled, or HCW sampling was anonymised and positive results could not be linked definitively with the suspected carrier.

The study necessarily reflects the circulating \textit{S. aureus} clones in the UK and the concerns of local infection control teams. The sampling frame is therefore enriched for MRSA and PVL-positive outbreaks and neonatal unit; despite this there is a wide representation of sequence types.

Despite the enhanced surveillance during each outbreak, there are inevitably missing transmission links, due to missed sampling, suppression from antimicrobial therapy, or delays in identifying contacts. One reason for missed samples may be the use of antibiograms as an initial screening tool for identifying putative outbreak isolates, as most investigating teams only collected isolates with identical or highly similar antimicrobial susceptibility profiles. However, in the six outbreaks where isolates were included with differing antibiograms, the core genomes were remarkably conserved. This is presumably
due to the ready loss/gain of mobile genetic elements (13), and shows that reliance on antibiograms may lead to samples being wrongly excluded.

The variability of mobile elements is also important for interpreting genetic distances. Recombination events such as gain/loss of a mobile element will introduce a large number of SNVs even though this represents a single genetic event. Current analysis tools which can accommodate this are computationally complex and, for large datasets, require sizable computing resources. A simpler approach is to exclude the “mobile-ome” from phylogenetic analyses and compare only the core genome, and the results above demonstrate that this is an acceptable strategy. Similarly, mapping to alternative reference strains (performed for six outbreaks) had minimal effect on SNV analysis and phylogeny, removing the need for identification of clonal complex or index case assembly prior to phylogenetic analysis. This streamlined approach brings WGS closer to routine use, as a readily deployable method with a minimal burden of computational time and bioinformatic expertise.

In conclusion, we have demonstrated how a WGS-based approach can be applied to S. aureus outbreak investigations. We have shown that current sampling strategies provide sufficient information to determine whether isolates belong to an outbreak, and that, rather than confounding the investigation, within-host diversity can be utilised to identify the possible involvement of a long-term carrier, potentially enhancing the infection control response. Combining this with directed multi-sampling of suspected LTCs (1) may be a cost-effective method of using WGS to ensure that, where HCWs are implicated, potentially career altering decisions are made using the best possible evidence.

**Methods**

**Comparison of within-host diversity in newly acquired and long-term carriage**

Eight participants were identified from a population study of S. aureus nasal carriage in adults attending general practices in Oxfordshire (14), in which participants had nasal swabs
taken at two-monthly intervals, with positive samples stored as mixed glycerol stocks taken by sweeping across multiple colonies on the primary plates to preserve the diversity of carried strains (11). The eight participants were negative for nasal carriage at recruitment and had consistently negative swabs for ≥6 months subsequently, before acquiring a strain which they carried for at least one year. The first and last positive samples for each individual were retrieved from the mixed glycerol stocks. Samples were plated on Columbia blood agar (CBA) and incubated overnight at 37°C. For each time-point, 8 individual colonies (12 for one individual, id=1218) were selected and sub-cultured to a further CBA plate and again incubated overnight at 37°C.

We also retrieved sequencing data from 13 participants previously investigated, for whom the approximate time of acquisition was unknown (9). Each of these had 8-12 individual colonies sequenced.

Collection of outbreak isolates and epidemiological data

19 outbreaks were purposively sampled in collaboration with the Public Health England (PHE) reference laboratory, representing a range of sequence-types and epidemiological settings, and including both MRSA and MSSA. One further outbreak was investigated in conjunction with Lausanne University Hospital, Switzerland (15, 16). Epidemiological information was obtained from each infection control team (specimen date, site, ward location and, where applicable, admission/discharge dates and previous screening results).

For each outbreak, additional background isolates were also included for comparison. We sequenced all isolates submitted to PHE as part of the outbreak investigation, including those identified as “non-outbreak” by routine typing, to estimate expected genetic diversity outwith the outbreak strain, and to ensure that the apparent outbreak strains were not part of an ongoing clonal expansion. We also included non-epidemiologically linked isolates.
matched for spa-type and/or MLST, to provide a comparison for expected within-spa
distances, and to provide an outgroup for phylogenetic analysis.

Isolates were retrieved from single colony frozen stocks held at the PHE reference
laboratory, Colindale, London, or at Lausanne University Hospital. We used only the first
isolate from each case, and included isolates both from clinical samples and screening
swabs.

**Extraction and sequencing**

DNA was extracted and sequenced as previously described (6) from a single colony sub-
cultured on CBA and incubated for 18-24hrs. Sequencing was performed using the Illumina
HiSeq or MiSeq platforms.

**Genome assembly and construction of phylogenetic trees**

For all outbreaks, reads were aligned using Stampy v1.0.17 to a standard reference genome
(MRSA252: GenBank NC_002952) (17). Six outbreaks were also mapped to clonal-complex
specific reference genomes obtained from in-house collections or GenBank. Single
nucleotide variants were identified across all mapped non-repetitive sites using SAMtools v
0.1.18 mpileup, with the extended base-alignment quality flag and masking of mobile genetic
elements. A consensus of ≥75%, and ≥5 reads, including one in each direction, was required
to support a SNV, and calls were required to be homozygous under a diploid model.
Maximum likelihood trees were estimated from the mapped whole genomes using PhyML
(18).

**Outbreak analysis and calculation of TMRCA**

The index case was defined as the earliest microbiologically confirmed case in each cluster.
Outbreak cases were defined as those sharing related PFGE pulsotypes (19) plus a definite
epidemiological link to the index or secondary cases (>24hr stay in same ward, or
household/classroom/similar community situation with prolonged contact e.g. childcare). For each outbreak case, the genetic distance in SNVs was calculated from the index case and the next nearest neighbour. If an isolate was more distant from the index case than the nearest spa/MLST-matched comparator, it was considered sporadic and excluded from further outbreak analysis.

We classified each outbreak according to the possibility of long-term carrier involvement (LTC, carrying for ≥6 weeks) as follows:

1) LTC not suspected: direct contact between cases, no history of pre-existing staphylococcal disease

2) evidence for a pre- or peri-outbreak LTC: either ≥1 case with prior history of recurrent staphylococcal disease, or non-overlapping hospital stays (ward case identified after a case-free interval, indicating a possible healthcare-worker carrier)

3) evidence of a post-outbreak LTC: ≥1 case with positive nasal swab >6 weeks after initial swab (indicating a propensity for long term carriage).

To evaluate the relationship between outbreak diversity and the likelihood of a long term carrier, we estimated time to most recent common ancestor (TMRCA) using BEAST v1.8.1 (20). We applied a simple HKY substitution model with constant population size and a standardized substitution rate of $3.3 \times 10^{-6}$ substitutions per genome per year (7) (see supplementary data). To control for differences in outbreak duration, outbreaks were censored at six months, and the (censored) outbreak duration subtracted from the calculated TMRCA to obtain a duration-adjusted TMRCA.

We compared SNV distances between isolates of identical pulsotype and those differing by one or more band. To determine whether there was an increase in genetic diversity
associated with acquisition of antimicrobial resistance, we also interrogated the predicted antibiograms as previously described (21).

Statistical analyses were performed using Stata v13.1. Mean pairwise differences were modelled using normal linear regression using robust standard errors to account for dependence within person/outbreak. The ability of TMRCA to differentiate between outbreaks with evidence for an LTC compared with outbreaks with no evidence for an LTC was evaluated using a receiver-operating-characteristic curve analysis.

The sequences reported in this paper have been deposited in the NCBI Sequence Read Archive under bioproject number PRJNA380544.

Acknowledgements
We are grateful to the staff of the PHE Staphylococcal reference service (especially Mrs Marjorie Ganner) for help in identifying and locating samples. We would also like to extend our thanks to the Infection Control teams at each site for permission to use isolates and epidemiological data, in particular the team at East Kent Hospitals NHS Foundation Trust.

Funding
This study is supported by the Health Innovation Challenge Fund (a parallel funding partnership between the Wellcome Trust [WT098615/Z/12/Z] and the Department of Health [grant HICF-T5-358]), and by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Healthcare Associated Infections and Antimicrobial Resistance at University of Oxford at in partnership with Public Health England (PHE) [grant number HPRU-2012-10041]. NCG is an MRC doctoral research fellow. The investigation of the Swiss outbreak was supported by a grant from the Swiss Foundation for Research (no. 31003A_150029). DWC and TEAP are NIHR Senior Investigators. The views expressed are
those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health, Wellcome Trust, the Medical Research Council or Public Health England.


### Table 1. Description of 20 outbreaks analysed by whole genome sequencing

<table>
<thead>
<tr>
<th>Outbreak Category</th>
<th>No of cases</th>
<th>Reason for outbreak investigation</th>
<th>MRSA or MSSA</th>
<th>Clonal complex</th>
<th>MLST</th>
<th>spa</th>
<th>Duration (days)</th>
<th>PFGE pulsortypes</th>
<th>Outbreak antibiograms</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Hospital - general ward</td>
<td>5</td>
<td>MRSA colonisation</td>
<td>MRSA</td>
<td>CC22</td>
<td>ST22</td>
<td>t032</td>
<td>367</td>
<td>All identical</td>
</tr>
<tr>
<td>B</td>
<td>Hospital - general ward</td>
<td>6</td>
<td>S. aureus wound infections</td>
<td>MSSA</td>
<td>CC8</td>
<td>ST2021</td>
<td>t008</td>
<td>412</td>
<td>All identical</td>
</tr>
<tr>
<td>C</td>
<td>Hospital - general ward</td>
<td>7</td>
<td>S. aureus wound infections</td>
<td>MRSA</td>
<td>CC8</td>
<td>ST239</td>
<td>t037</td>
<td>98</td>
<td>All identical</td>
</tr>
<tr>
<td>D</td>
<td>Hospital - general ward</td>
<td>17</td>
<td>MRSA colonisation</td>
<td>MRSA</td>
<td>CC8</td>
<td>ST8</td>
<td>t008</td>
<td>88</td>
<td>All identical</td>
</tr>
<tr>
<td>E</td>
<td>Hospital – surgical unit</td>
<td>8</td>
<td>S. aureus wound infections</td>
<td>MRSA</td>
<td>CC22</td>
<td>ST22</td>
<td>t022</td>
<td>18</td>
<td>All identical</td>
</tr>
<tr>
<td>F</td>
<td>Hospital - multiple wards</td>
<td>50</td>
<td>MRSA colonisation</td>
<td>MRSA</td>
<td>CC5</td>
<td>ST228</td>
<td>t041</td>
<td>122</td>
<td>2 pulsortypes</td>
</tr>
<tr>
<td>G</td>
<td>Hospital – surgical unit</td>
<td>187</td>
<td>MRSA colonisation</td>
<td>MRSA</td>
<td>CC5</td>
<td>ST228</td>
<td>t041</td>
<td>454</td>
<td>All identical</td>
</tr>
<tr>
<td>H</td>
<td>Hospital – maternity unit</td>
<td>6</td>
<td>PVL-related SSTIs</td>
<td>MRSA</td>
<td>CC1</td>
<td>ST772</td>
<td>t657</td>
<td>70</td>
<td>All identical</td>
</tr>
<tr>
<td>I</td>
<td>Hospital – maternity unit</td>
<td>9</td>
<td>Scalded skin syndrome</td>
<td>MSSA</td>
<td>CC15</td>
<td>ST2434</td>
<td>t346</td>
<td>70</td>
<td>All identical</td>
</tr>
<tr>
<td>J</td>
<td>Hospital – neonatal unit</td>
<td>3</td>
<td>MRSA colonisation</td>
<td>MRSA</td>
<td>CC59</td>
<td>ST59</td>
<td>t216</td>
<td>8</td>
<td>All identical</td>
</tr>
<tr>
<td>K</td>
<td>Hospital – neonatal unit</td>
<td>4</td>
<td>MRSA colonisation</td>
<td>MRSA</td>
<td>CC22</td>
<td>ST22</td>
<td>t5692</td>
<td>43</td>
<td>All identical</td>
</tr>
<tr>
<td>L</td>
<td>Hospital – neonatal unit</td>
<td>6</td>
<td>MRSA colonisation</td>
<td>MRSA</td>
<td>CC30</td>
<td>ST30</td>
<td>t019</td>
<td>57</td>
<td>All identical</td>
</tr>
<tr>
<td>M</td>
<td>Hospital – neonatal unit</td>
<td>8</td>
<td>MRSA bacteraemia</td>
<td>MRSA</td>
<td>CC30</td>
<td>ST86</td>
<td>t5973</td>
<td>65</td>
<td>All identical</td>
</tr>
<tr>
<td>N</td>
<td>Hospital – neonatal unit</td>
<td>41</td>
<td>MRSA colonisation</td>
<td>MRSA</td>
<td>CC22</td>
<td>ST22</td>
<td>t5892</td>
<td>1526</td>
<td>All identical</td>
</tr>
<tr>
<td>O</td>
<td>Household</td>
<td>3</td>
<td>PVL-related SSTIs</td>
<td>MRSA</td>
<td>CC30</td>
<td>ST30</td>
<td>t019</td>
<td>8</td>
<td>All identical</td>
</tr>
<tr>
<td>P</td>
<td>Household</td>
<td>4</td>
<td>PVL-related SSTIs</td>
<td>MRSA</td>
<td>CC30</td>
<td>ST30</td>
<td>t019</td>
<td>20</td>
<td>3 pulsortypes</td>
</tr>
<tr>
<td>Q</td>
<td>Household</td>
<td>5</td>
<td>PVL-related SSTIs</td>
<td>MRSA</td>
<td>CC30</td>
<td>ST30</td>
<td>t019</td>
<td>195</td>
<td>2 pulsortypes</td>
</tr>
<tr>
<td>R</td>
<td>Household</td>
<td>8</td>
<td>PVL-related SSTIs</td>
<td>MRSA</td>
<td>CC30</td>
<td>ST30</td>
<td>t019</td>
<td>44</td>
<td>All identical</td>
</tr>
<tr>
<td>S</td>
<td>Nursing home</td>
<td>9</td>
<td>PVL-related SSTIs</td>
<td>MRSA</td>
<td>CC30</td>
<td>ST30</td>
<td>t019</td>
<td>298</td>
<td>2 pulsortypes</td>
</tr>
<tr>
<td>T</td>
<td>School</td>
<td>5</td>
<td>PVL-related SSTIs</td>
<td>MSSA</td>
<td>CC121</td>
<td>ST121</td>
<td>t645</td>
<td>74</td>
<td>All identical</td>
</tr>
</tbody>
</table>

PVL: Panton-Valentine Leukocidin; MLST: multi-locus sequence-type
Figure 1. All pairwise differences between early (<2 months since acquisition) and late (>12 months since acquisition) nasal swab samples from 7 patients with previous negative nasal swabs. Included for comparison are samples from patients positive at entry to the study (time of acquisition unknown).
Table 2: Long term carrier category, duration-adjusted TMRCA and SNV range for outbreaks investigated using WGS

<table>
<thead>
<tr>
<th>Outbreak</th>
<th>Long term carrier category</th>
<th>Duration-adjusted TMRCA, days (95% highest posterior density interval)</th>
<th>Range of distances between all isolates in cluster, SNVs</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Indirect ward contact</td>
<td>285 (134-445)</td>
<td>0-19</td>
</tr>
<tr>
<td>B</td>
<td>Indirect ward contact</td>
<td>515 (394-646)</td>
<td>0-24</td>
</tr>
<tr>
<td>C</td>
<td>Direct ward contact</td>
<td>103 (61-228)</td>
<td>0-9</td>
</tr>
<tr>
<td>D</td>
<td>Direct ward contact</td>
<td>78 (21-153)</td>
<td>0-10</td>
</tr>
<tr>
<td>E</td>
<td>Post outbreak LTC</td>
<td>96 (40-165)</td>
<td>0-6</td>
</tr>
<tr>
<td>F</td>
<td>Direct ward contact</td>
<td>86 (36-160)</td>
<td>0-5</td>
</tr>
<tr>
<td>G</td>
<td>Indirect ward contact</td>
<td>394 (259-539)</td>
<td>0-46</td>
</tr>
<tr>
<td>H</td>
<td>Direct ward contact</td>
<td>46 (5-95)</td>
<td>0-4</td>
</tr>
<tr>
<td>I</td>
<td>Post outbreak LTC</td>
<td>30 (1-69)</td>
<td>0-5</td>
</tr>
<tr>
<td>J</td>
<td>Post outbreak LTC</td>
<td>161 (61-271)</td>
<td>0-9</td>
</tr>
<tr>
<td>K</td>
<td>Direct ward contact</td>
<td>31 (0-82)</td>
<td>1-4</td>
</tr>
<tr>
<td>L</td>
<td>Post outbreak LTC</td>
<td>216 (134-314)</td>
<td>0-8</td>
</tr>
<tr>
<td>M</td>
<td>Direct ward contact</td>
<td>54 (9-107)</td>
<td>0-4</td>
</tr>
<tr>
<td>N</td>
<td>Indirect ward contact</td>
<td>156 (50-275)</td>
<td>0-36</td>
</tr>
<tr>
<td>O</td>
<td>Pre-outbreak LTC</td>
<td>378 (240-531)</td>
<td>9-25</td>
</tr>
<tr>
<td>P</td>
<td>Pre-outbreak LTC</td>
<td>54 (12-124)</td>
<td>1-11</td>
</tr>
<tr>
<td>Q</td>
<td>Pre-outbreak LTC</td>
<td>204 (108-306)</td>
<td>1-13</td>
</tr>
<tr>
<td>R</td>
<td>Household</td>
<td>29 (0-73)</td>
<td>1-2</td>
</tr>
<tr>
<td>S</td>
<td>Indirect ward contact</td>
<td>431 (269-599)</td>
<td>3-32</td>
</tr>
<tr>
<td>T</td>
<td>School</td>
<td>12 (0-40)</td>
<td>0-2</td>
</tr>
</tbody>
</table>
Figure 2. Duration-adjusted TMRCA for outbreaks with 1) no evidence of a long term carrier (direct contacts between all cases); 2) likely LTC (indirect ward contacts or pre-outbreak LTC); 3) LTC unclear / possible (evidence of a post-outbreak LTC)
Figure 3. Pairwise SNV differences for all pairs within an outbreak, where isolates had differing antibiograms (a) or differing PFGE pulsotypes (b).