

## INFECTIOUS DISEASE

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

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

Copyright © 2017  
The Authors, some  
rights reserved;  
exclusive licensee  
American Association  
for the Advancement  
of Science. No claim  
to original U.S.  
Government Works

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

## INTRODUCTION

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

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

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

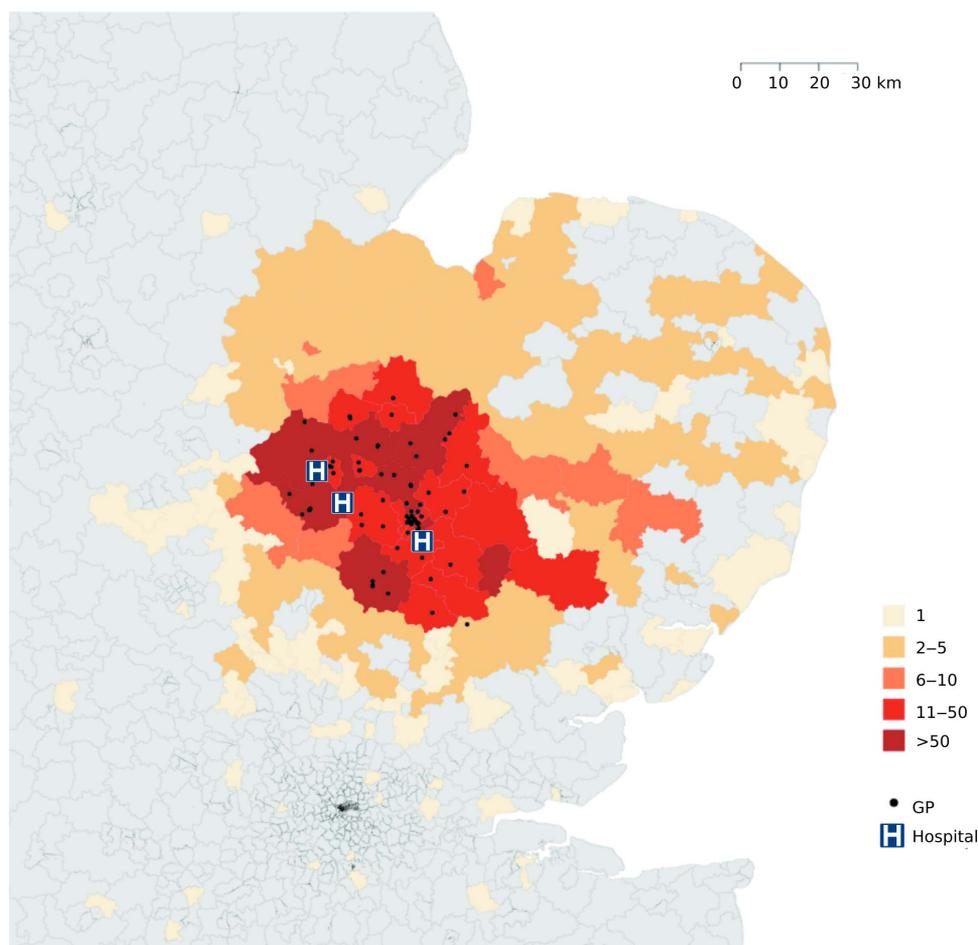
## RESULTS

### Study participants and MRSA isolates

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

<sup>1</sup>London School of Hygiene and Tropical Medicine, London, UK. <sup>2</sup>University of Cambridge, Cambridge, UK. <sup>3</sup>Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK. <sup>4</sup>Wellcome Trust Sanger Institute, Cambridge, UK. <sup>5</sup>Public Health England, London, UK. <sup>6</sup>Papworth Hospital NHS Foundation Trust, Cambridge, UK.

\*Corresponding author. Email: francesc.coll@lshtm.ac.uk (F.C.); sharon.peacock@lshtm.ac.uk (S.J.P.)



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

### Integration of genomic and epidemiological data

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

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

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

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

### Evidence of MRSA transmission in the community

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

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

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

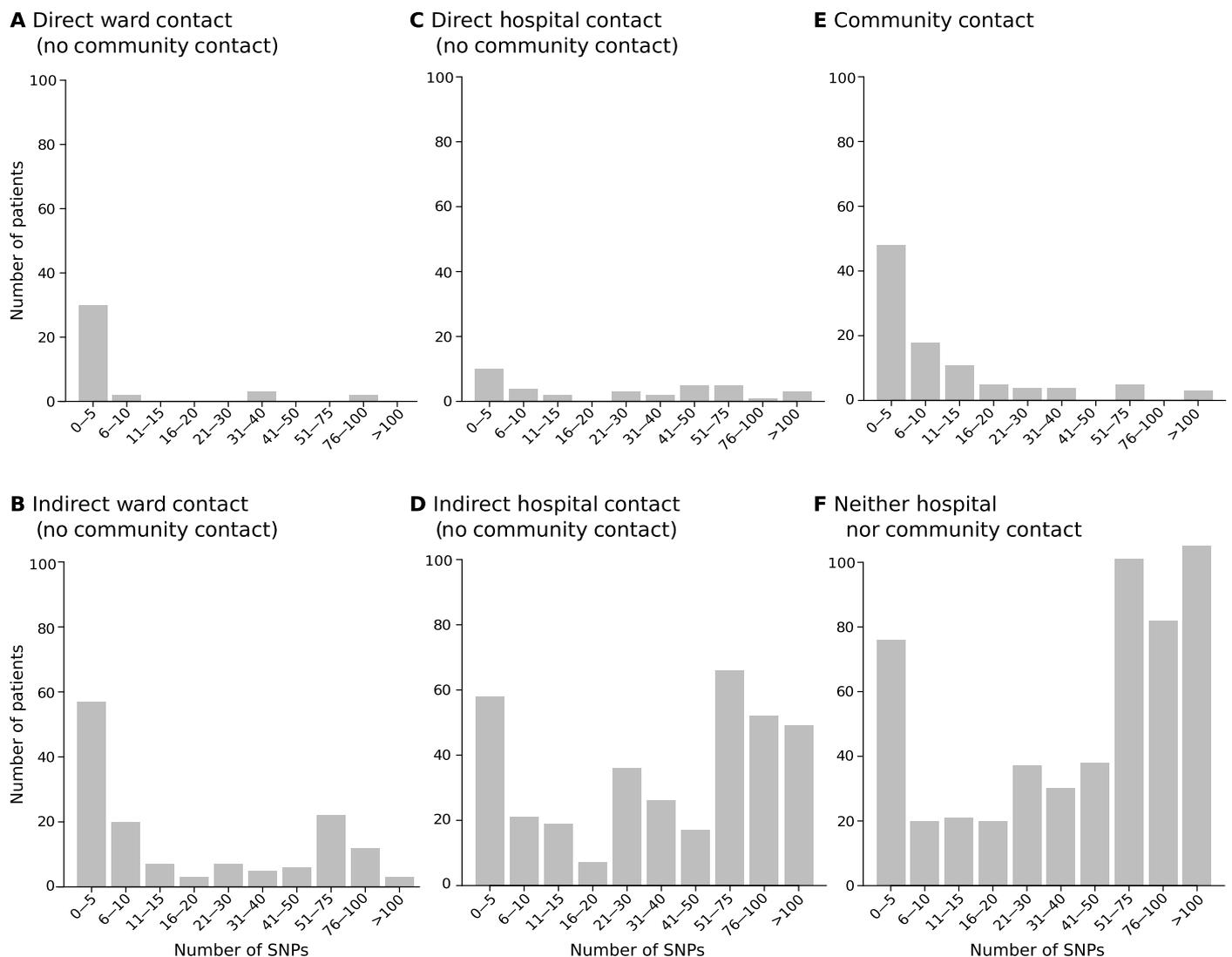
8 long-term care facilities (22 cases) (Table 1). A pictorial representation of exemplars of transmission at a GP practice, long-term care facility, and household is shown in fig. S4 (A to C).

**Evidence of MRSA transmission in hospitals**

More than half of cases with epidemiological and bacterial genomic links (371 of 598, 62%) resided in transmission clusters with hospital

contacts, of which 255 cases had ward contacts. The 371 cases were resolved into 118 different clusters each involving between 2 and 44 individuals (Table 1). We narrowed down further investigation to those clusters that contained five or more patients (nine clusters; see table S3 for details) and evaluated these for instances of direct ward contact (same ward, overlapping admission dates) or indirect ward contact (same ward, no overlap in admission dates). Where available,

Downloaded from https://www.science.org on December 17, 2023



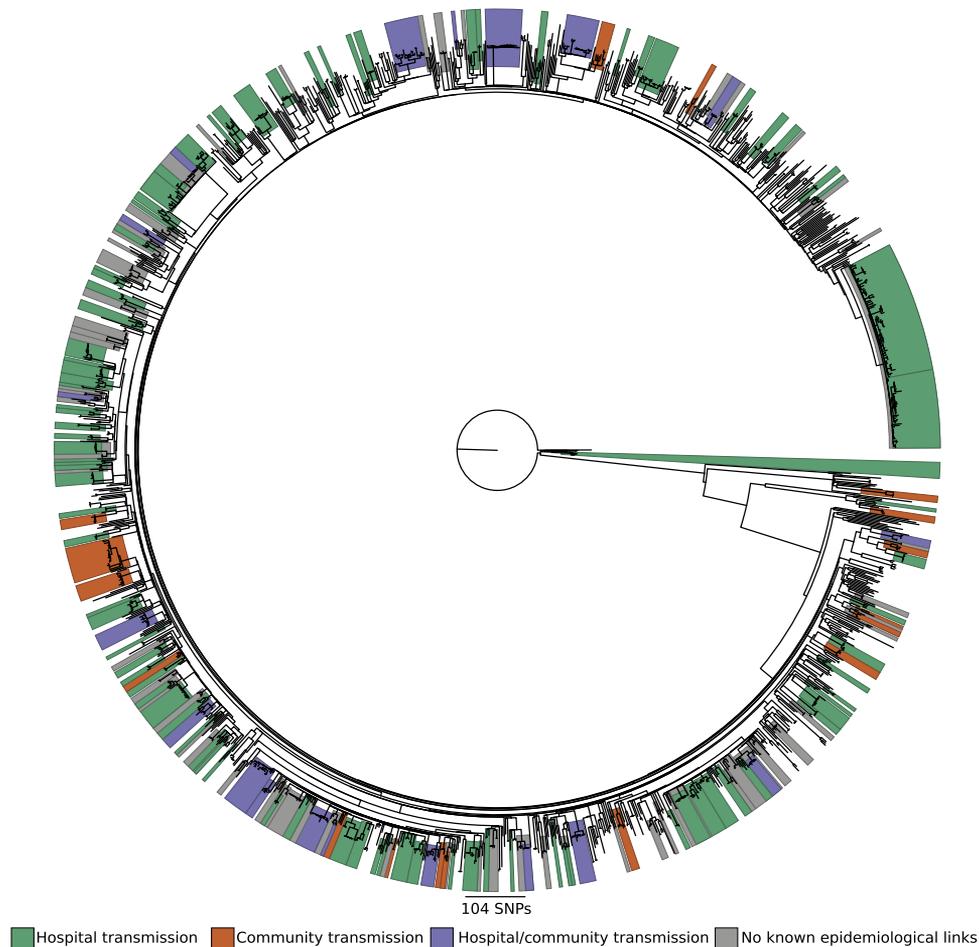
**Fig. 2. Pairwise comparison between MRSA relatedness and type of patient contact.** For each case, the most closely related MRSA isolate from another case was identified, and the epidemiological contact of each case pair was defined. The number of cases in each epidemiological category is shown as a function of the genetic distance (difference in the number of SNPs in the core genome). (A to D) Genetic distance distribution for cases with hospital contacts alone. Direct contact refers to a link in the same time and place (ward or hospital). Indirect contact refers to a link in the same place but different time. (E) Community contacts (shared residential postcodes or GP practice). (F) Cases with neither hospital nor community contacts. Only cases with MRSA isolates from CCs found in at least one other patient in the population are shown ( $n = 1459$ ).

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

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

#### MRSA transmission at the hospital-community interface

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



**Fig. 3. Transmission clusters color-coded on the CC22 phylogeny.** Maximum likelihood tree generated from 34,600 SNP sites in the core genome is shown for 1667 CC22 isolates. Colors refer to the type of epidemiological links in clusters of genetically related isolates (maximum 50 SNPs) from multiple cases.

## DISCUSSION

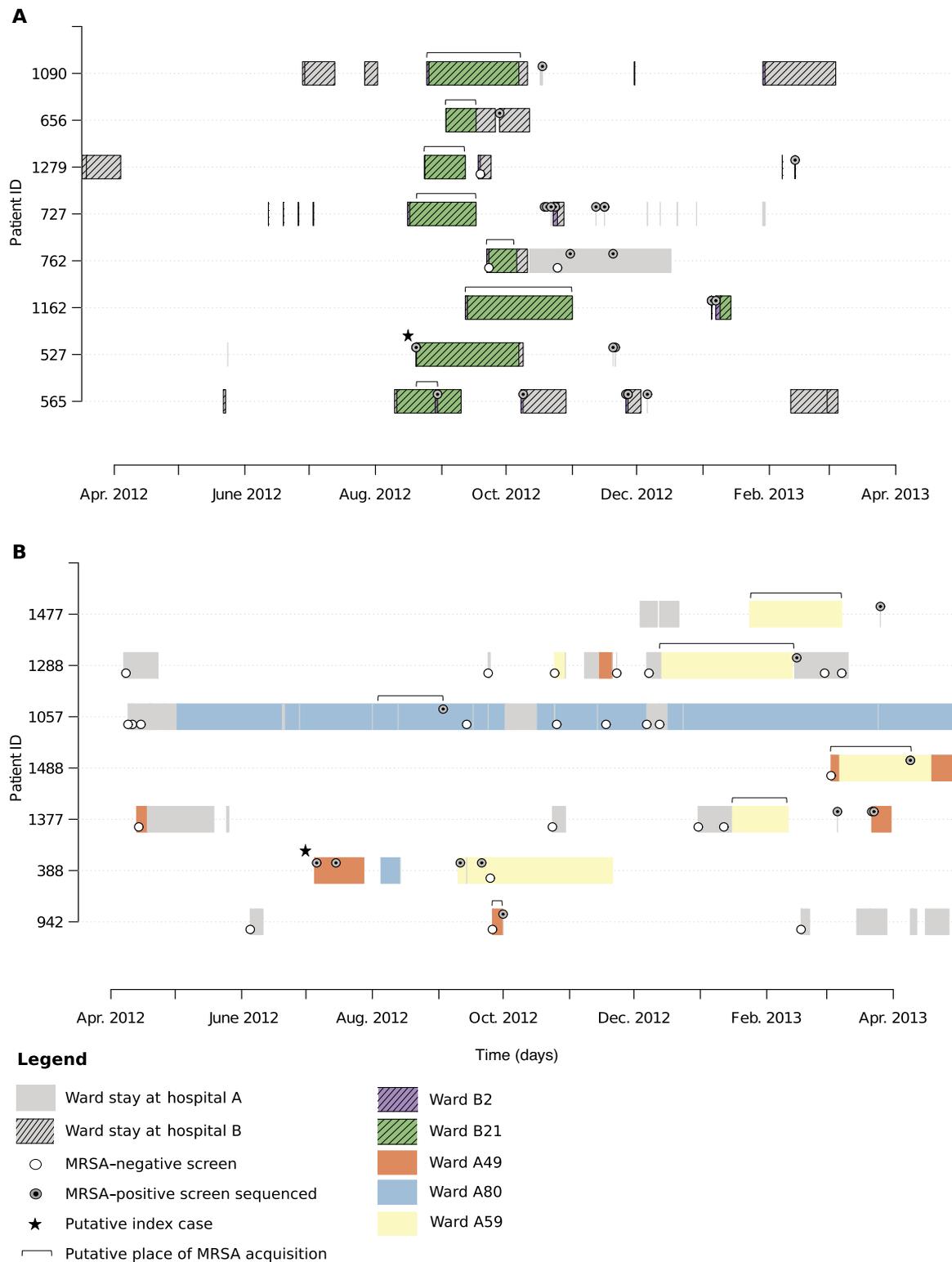
Our findings have important implications for infection control policy and practice. MRSA transmission in our study population was not attributable to large nosocomial outbreaks but resulted from the cumulative effect of numerous clinically unrecognized episodes. We detected 173 separate genetic clusters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

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

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

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

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



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

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

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

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

## MATERIALS AND METHODS

### Study design

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

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

### DNA sequencing and genomic analyses

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

### Epidemiological analysis

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

### Identification of putative MRSA transmission

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

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

## SUPPLEMENTARY MATERIALS

[www.sciencetranslationalmedicine.org/cgi/content/full/9/413/eaak9745/DC1](http://www.sciencetranslationalmedicine.org/cgi/content/full/9/413/eaak9745/DC1)  
Materials and Methods

Fig. S1. Number of isolates sequenced per patient.

Fig. S2. Flowchart summarizing data types and analyses.

Fig. S3. Integration of genomic and epidemiological data to identify transmission clusters.

Fig. S4. Six examples of transmission clusters in different settings.

Fig. S5. Number of heterozygous sites in the core genome per isolate.

Fig. S6. Within-host diversity over time and at a single time point.

Table S1. Proportion of isolates in different CCs.

Table S2. Frequency of epidemiological contacts among genetically unrelated cases.

Table S3. Epidemiological classification of transmission clusters containing five or more cases.

Data file S1. Accession numbers.

References (22–26)

## REFERENCES AND NOTES

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

**Acknowledgments:** We thank H. Brodrick, K. Judge, H. Giramahoro, and M. Blackman-Northwood for technical assistance; L. Mlemba for clinical data collection; the Wellcome Trust Sanger Institute Core Sequencing and Pathogen Informatics Groups; and D. Harris for assisting in submitting sequence data to public databases. **Funding:** This work was supported by grants from the UK Clinical Research Collaboration Translational Infection Research Initiative and the Medical Research Council (grant no. G1000803) with contributions to the grant from the

Biotechnology and Biological Sciences Research Council, the National Institute for Health Research (NIHR) on behalf of the Department of Health, and the Chief Scientist Office of the Scottish Government Health Directorate (to S.J.P.); by a Hospital Infection Society Major Research Grant; by Wellcome Trust grant no. 098051 awarded to the Wellcome Trust Sanger Institute; and by Wellcome Trust 201344/Z/16/Z awarded to F.C. M.S.T. is a Wellcome Trust Clinical PhD fellow. M.E.T. is a Clinician Scientist Fellow, supported by the Academy of Medical Sciences and the Health Foundation and by the NIHR Cambridge Biomedical Research Centre.

**Author contributions:** M.E.T. and S.J.P. designed the study, wrote the study protocol and case record forms, obtained ethical and research and development approvals for the study, and supervised the data collection. N.M.B., A.R.M.K., and B.P. were responsible for isolating and identifying MRSA in the diagnostic microbiology laboratory and provided expert opinion relating to infection control. F.C. undertook the epidemiological and bioinformatic analyses with contributions from E.M.H., M.S.T., and S.R. B.B. and K.E.R. conducted the laboratory work. J.P. supervised the genomic sequencing. F.C. and S.J.P. wrote the first draft of the manuscript. S.J.P. supervised and managed the study. All authors had access to the data and read, contributed, and approved the final manuscript. **Competing interests:** N.M.B. is on the advisory

board for Discuva Ltd. S.J.P. and J.P. are paid consultants for Specific Technologies.

All other authors declare that they have no competing interests. **Data and materials availability:** The whole-genome sequences from this study have been deposited in the European Nucleotide Archive under study accession no. PRJEB3174. Run accession numbers are listed in data file S1.

Submitted 23 September 2016

Resubmitted 24 March 2017

Accepted 10 July 2017

Published 25 October 2017

10.1126/scitranslmed.aak9745

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

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

Francesc Coll, Ewan M. Harrison, Michelle S. Toleman, Sandra Reuter, Kathy E. Raven, Beth Blane, Beverley Palmer, A. Ruth M. Kappeler, Nicholas M. Brown, M. Estée Török, Julian Parkhill, and Sharon J. Peacock

*Sci. Transl. Med.* **9** (413), eaak9745. DOI: 10.1126/scitranslmed.aak9745

### On the trail of MRSA

Genome sequencing of methicillin-resistant *Staphylococcus aureus* (MRSA) has been successfully applied to investigate suspected outbreaks. Coll *et al.* now extend its application to the genomic surveillance of MRSA in samples from 1465 people identified over a 12-month period by a diagnostic laboratory in the East of England. This analysis identified 173 putative outbreaks involving 598 patients and included hospital outbreaks, those spanning the hospital and community, and community outbreaks among people registered with the same medical practice or living in the same household or long-term care facility. This study illustrates that sequencing is a powerful tool that could be used to identify infectious disease outbreaks as they happen.

### View the article online

<https://www.science.org/doi/10.1126/scitranslmed.aak9745>

### Permissions

<https://www.science.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of service](#)

---

*Science Translational Medicine* (ISSN 1946-6242) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Translational Medicine* is a registered trademark of AAAS.

Copyright © 2017 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works