THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Evolutionary Dynamics of Pandemic Methicillin-Sensitive Staphylococcus aureus ST398 and Its International Spread via Routes of Human Migration

OPEN ACCESS

Evolutionary Dynamics of Pandemic Methicillin-Sensitive *Staphylococcus aureus* ST398 and Its International Spread via Routes of Human Migration

Anne-Catrin Uhlemann,[a] Paul R. McAdam,[b,c] Sean B. Sullivan,[a] Justin R. Knox,[a] Hossein Khiabanian,[d,e] Raul Rabadan,[e] Peter R. Davies,[f] J. Ross Fitzgerald,[b] Franklin D. Lowy[a,g]

Department of Medicine, Columbia University Medical Center, New York, New York, USA[a]; The Roslin Institute, University of Edinburgh, Edinburgh, Scotland, United Kingdom[b]; Department of Biochemistry and Molecular Biology, University of Melbourne, Parkville, Victoria, Australia[c]; Rutgers Cancer Institute of New Jersey, Rutgers University, New Brunswick, New Jersey, USA[d]; Department of Systems Biology, Columbia University, New York, New York, USA[e]; Department of Veterinary Population Medicine, University of Minnesota, St. Paul, Minnesota, USA[f]; Department of Pathology, Columbia University Medical Center, New York, New York, USA[g]

ABSTRACT Methicillin-susceptible *Staphylococcus aureus* (MSSA) accounts for the majority of *S. aureus* infections globally, and yet surprisingly little is known about its clonal evolution. We applied comparative whole-genome sequencing (WGS) analyses to epidemiologically and geographically diverse ST398-MSSA, a pandemic lineage affecting both humans and livestock. Bayesian phylogenetic analysis predicted divergence of human-associated ST398-MSSA ~40 years ago. Isolates from Midwestern pigs and veterinarians differed substantially from those in New York City (NYC). Pig ST398 strains contained a large region of recombination representing imports from multiple sequence types (STs). Phylogeographic analyses supported the spread of ST398-MSSA along local cultural and migratory links between parts of the Caribbean, North America, and France, respectively. Applying pairwise single-nucleotide polymorphism (SNP) distances as a measure of genetic relatedness between isolates, we observed that ST398 not only clustered in households but also frequently extended across local social networks. Isolates collected from environmental surfaces reflected the full diversity of colonizing individuals, highlighting their potentially critical role as reservoirs for transmission and diversification. Strikingly, we observed high within-host SNP variability compared to our previous studies on the dominant methicillin-resistant *Staphylococcus aureus* (MRSA) clone USA300. Our data indicate that the dynamics of colonization, persistence, and transmission differ substantially between USA300-MRSA and ST398-MSSA. Taken together, our study reveals local and international routes of transmission for a major MSSA clone, indicating key impacts of recombination and mutation on genetic diversification and highlighting important ecological differences from epidemic USA300. Our study demonstrates extensive local and international routes of transmission for a major MSSA clone despite the lack of substantial antibiotic resistance.

IMPORTANCE Unlike methicillin-resistant *Staphylococcus aureus* (MRSA), surprisingly little is known about the clonal evolution of methicillin-susceptible *S. aureus* (MSSA), although these strains account for the majority of *S. aureus* infections. To better understand how MSSA spreads and becomes established in communities, we applied comparative bacterial whole-genome sequencing to pandemic ST398-MSSA, a clone of clinical importance affecting humans and livestock in different geographic regions. Phylogeographic analyses identified that ST398-MSSA spread along local cultural and migratory links between parts of the Caribbean, North America, and

Address correspondence to Anne-Catrin Uhlemann, au2110@columbia.edu.

A.-C.U., P.R.M., J.R.F., and F.D.L. contributed equally to this work.

France, respectively. We observed high within-host SNP variability compared to our previous studies on the dominant MRSA clone USA300. Our data indicate that the dynamics of colonization, persistence, and transmission differ substantially between USA300 MRSA and ST398 MSSA.

*S*taphylococcus aureus causes the majority of skin and soft tissue infections (SSTIs) and is also frequently associated with invasive disease, such as bloodstream infections or pneumonia (1). Over the past few decades, novel methicillin-resistant *S. aureus* (MRSA) genotypes have emerged due to the acquisition of variants of the staphylococcal cassette chromosome *mec* (SCC*mec*) element. Only a minority of emergent MRSA clones have disseminated globally and have caused epidemic waves of health care-associated (HA), community-associated (CA), or, more recently, livestock-associated (LA) *S. aureus* infections.

The advent of high-resolution whole-genome sequencing (WGS) has allowed the reconstruction of the evolution of several major MRSA clones such as multilocus sequence types (STs) ST239, ST22, ST30, ST80, and ST8/USA300 by sequencing of large data sets (2–7). Collectively, these studies have documented how specific antibiotic selective pressures, as well as interactions with human hosts, have shaped the recent evolutionary history of these MRSA clones. While these studies have included a selected number of methicillin-susceptible *S. aureus* (MSSA) isolates (7), the emergence, evolution, and transmission of major MSSA lineages have been largely disregarded. MSSA, however, remains of considerable public health importance as it accounts for the majority of health care- and community-based *S. aureus* infections throughout the world (8). In a possible reversal of past trends, it has been suggested that MSSA now accounts for higher numbers of HA-*S. aureus* (HA-SA) infections than does MRSA (9). Although the clonal backgrounds of MSSAs are highly diverse, a number of dominant pandemic MSSA clones have been identified (8, 10, 11). Major MSSA lineages overlap the dominant circulating MRSA clones and include ST30, ST5, ST8, and, more recently, ST398 (12). ST398 was first described as both MSSA and MRSA among pig farmers in France (13, 14). Since the original description, MRSA ST398 has quickly spread among pigs and other types of livestock and has been associated with MRSA infections in farmers (15, 16). MRSA ST398 strains are thought to spread only infrequently beyond the immediate animal and farm contacts (17–19).

In parallel, animal-independent human colonization and infections with ST398 MSSA have emerged and have now been encountered worldwide, including in several European countries (14, 20, 21), the Caribbean (22), and the northeastern United States (23, 24). The prevalence of ST398 MSSA appears to be particularly high in parts of China, where this clone accounts for almost 20% of skin and soft tissue infections (SSTIs) (25). This lineage has been notably absent from other regions such as large parts of the United States (11), although it has sporadically been reported in pigs and veterinarians in the Midwest (26). ST398 MSSA strains are encountered as both CA and HA pathogens (27, 28). Remarkably, a number of studies from such diverse settings as community households, a jail holding tank, and an intensive care unit have suggested that ST398 MSSA might be uniquely transmissible between humans (24, 27, 29). In addition to the distinct epidemiology, important genomic differences between human MSSA and LA-MRSA ST398 include a large repertoire of mobile genetic elements (MGEs) (30) and tetracycline resistance and lack of a unique SA3*int* β-hemolysin converting phage (21, 24, 31). In contrast, most of the human-associated ST398 isolates harbor resistance to macrolides and are *spa* type t571. WGS analyses indicated a possible human origin for LA-ST398, followed by the emergence of methicillin resistance driven by antibiotic pressure in animal feeds (31). A more recent quantitative time-scaled phylogeny, however, indicated that both human MSSA and LA-CC398 emerged in parallel around 1970 (32).

To further delineate the basis of the wide geographic distribution of ST398 MSSA, we investigated the macro- and microevolution of this clone by comparative whole-

genome sequencing of epidemiologically linked isolates from social networks in New York City, farms in the midwestern United States, and geographically distinct Caribbean islands. Our analysis provides insights into the recent adaptation and international spread of this highly successful MSSA clone.

RESULTS

Phylogenetic reconstruction of human ST398 MSSA isolates. In order to determine the evolutionary history of human ST398 MSSA, we determined full-genome sequences of 288 *spa*-clonal complex (CC) t571/ST398 isolates (Fig. 1; see also Table S1 in the supplemental material). These isolates were selected from community-based studies on the transmission of *S. aureus* in Northern Manhattan (NM) and the Bronx (22, 24, 28, 33) and from t571-MSSA-colonized veterinarians and pigs in Minnesota. To increase the genetic, temporal, and geographic diversity in the sample set for comparative sequence analyses, we also included previously published ST398 MSSA and MRSA sequences from international HA- and LA-CC398 collections (31).

Mapping to the ST398 reference genome and subsequent phylogenetic reconstruction revealed a distinct long branch length for 12 of the 14 isolates from veterinarians and pigs. These samples harbored a unique ~250-kb region at the *ori* compared to the other isolates (Fig. 1A). Comparison of this region and previously published reference genomes revealed that this sequence is of mosaic origin, suggestive of multiple large-scale recombination events (34) (Fig. 1B). The remainder of the sequence in these 12 isolates was consistent with an ST398 clonal background.

After exclusion of these 12 isolates, we removed mobile genetic elements (MGEs), predicted recombinant regions from the alignment, and identified 9,843 single nucleotide polymorphisms (SNPs) in comparison with the ST398 MSSA reference sequence (24). These "core" SNPs were used to construct a maximum-likelihood phylogenetic tree (Fig. 2A).

Mapping the geographic and host origin of isolates onto the phylogenetic tree showed clades with significant associations with both host and location ($P < 0.01$ for both traits, Slatkin-Maddison test; see Fig. S1 in the supplemental material). Isolates from the NM collection were relatively closely related and clustered separately from previously published LA-MRSA and most human CC398 strains from Europe (31) (Fig. 2A). The two isolates without evidence of large-scale recombination, collected from veterinarians in Minnesota, clustered with LA-CC398 isolates rather than with NM ST398 isolates.

Dominican isolates were interspersed with the majority of the isolates from the largely Dominican NM neighborhood. ST398 isolates from Martinique formed a distinct clade that also contained previously reported sequence from France (31) and several NM isolates (Fig. 2A). Within the NM clades, ST398 isolates from infections were interspersed with colonizing isolates (Fig. 2B). Importantly, none of the previous ST398 isolates from hospital-associated infections (28), which had been thought to represent a potential nosocomial ST398 outbreak, were closely related.

Time scale of the emergence of human CC398 in Northern Manhattan and phylogeographic reconstruction. To further reconstruct the evolution and geographic spread of the human-associated ST398 lineage, we selected isolates that showed evidence for clocklike evolution ($n = 119$) for Bayesian phylogeny. These included all geographically diverse isolates from Martinique and the Dominican Republic (DR) ($n = 14$) and Minnesota ($n = 2$) and from NM infectious isolates ($n = 69$) and colonizing isolates ($n = 11$) and a limited number of network isolates ($n = 23$). We estimated a substitution rate at 1.49×10^{-6} substitutions per site per year in the BEAST analysis (95% highest posterior density [HPD] interval = 9.44×10^{-7} to 1.97×10^{-6}). This rate is largely consistent with rates that have been previously reported for other *S. aureus* clones and for the CC398 lineage (5, 32) and is consistent with the estimate obtained from a root-to-tip regression of the full data set (1.47×10^{-6} substitutions per site per year) (see Fig. S2 in the supplemental material). The median time of clonal emergence was estimated around 36.6 years ago (95% HPD, 24.3 to 54.6) with a most

Uhlemann et al.
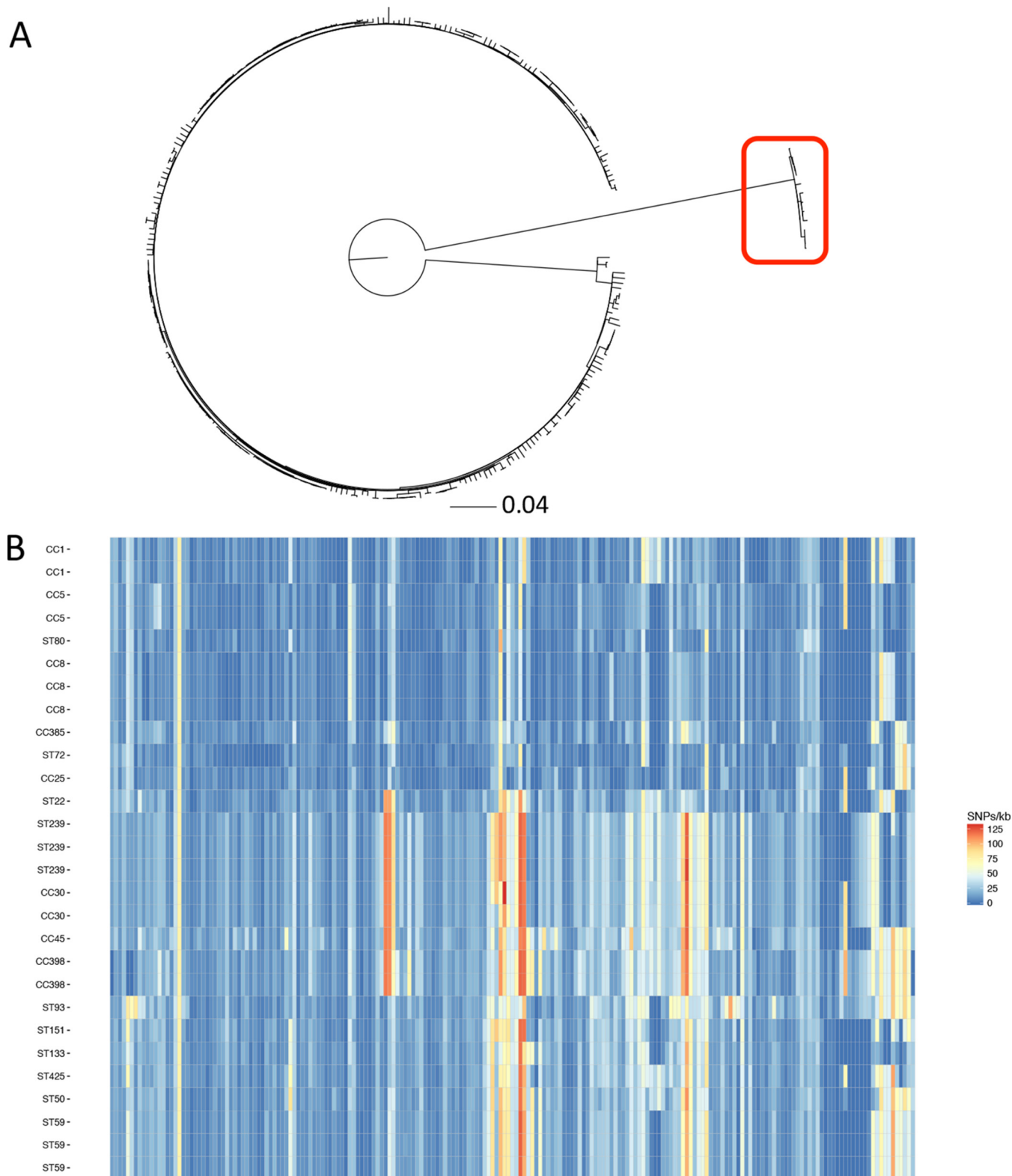
FIG 1 Phylogeny of ST398 reveals large-region recombination. (A) Maximum likelihood phylogenetic tree of ST398 isolates, rooted by midpoint, indicates that a subset of isolates (12 of the 14 isolates from veterinarians and pigs) is distinct. (B) Heat map of alignment of the ~250-kb region of recombination at *ori* indicates a mosaic origin from multiple clonal complexes (CC385, CC25, CC1, CC5, CC97, and CC8) rather than from a single clone. Single-nucleotide polymorphism density across the large recombinant region spanning origin of replication in the veterinary/animal strains with recombinant region (reference genome coordinates start at 2557000 and end at 58000). Each cell represents a block of 1,000 bp, and shading indicates density of SNPs in that region, with red indicating greater divergence. A representative selection of samples from both animals and humans was included.

FIG 2 ST398 phylogeny and emergence of ST398. (A) Maximum likelihood phylogenetic tree of ST398 indicates that isolates from Northern Manhattan form a distinct clade of human MSSA isolates. The outer ring and branches are colored by host origin, the middle ring indicates geographic region, and the inner ring shows the presence or absence of the SCC*mec* element. (B) ML phyloge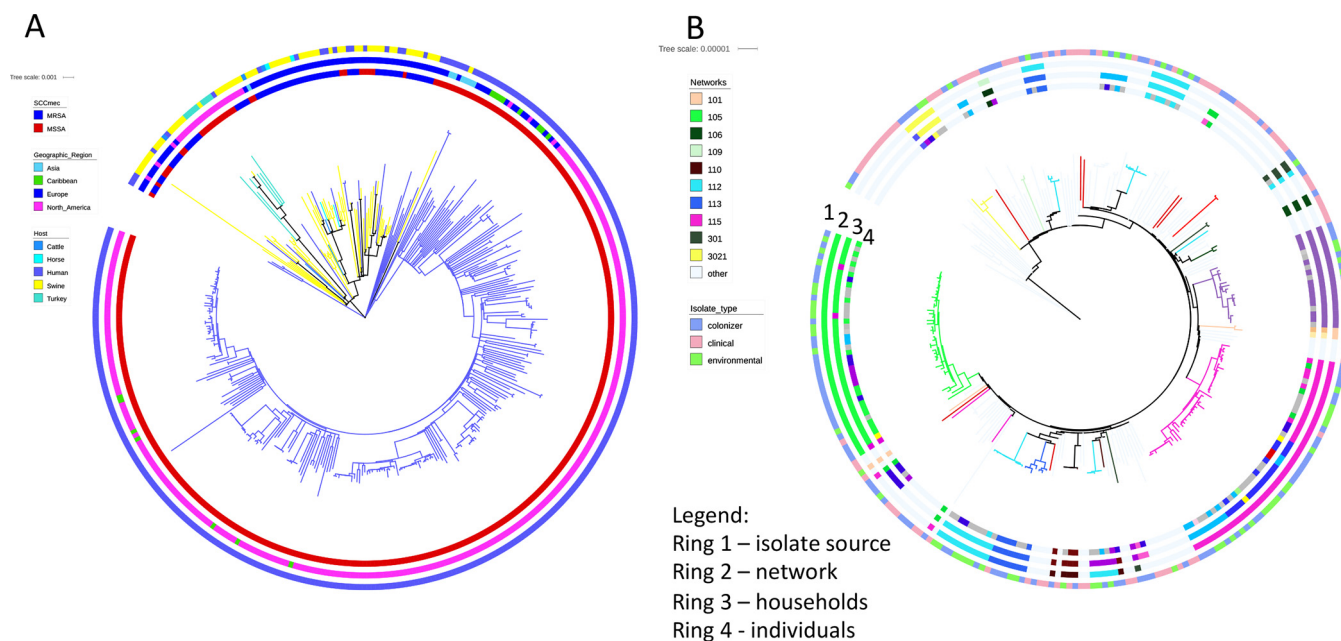netic tree indicates mapping of isolate source (outer ring 1) and social networks (ring 2 and colored branches). Ring 3 shows households within networks (index household color matches network). Inner ring 4 highlights the overlap of isolates collected from different individuals and environmental surfaces within network households (index color matches network color). Gray indicates environmental isolates. Light blue in circles 2 to 4 represents nonnetwork isolates.

recent common ancestor (MRCA) of the ST398 MSSA lineage in 1976 (Fig. 3). These findings are in accordance with a prior study including CC398 LA and HA clades that estimated their divergence around 1971 (32).

This scaled phylogeny also indicates that ST398 isolates from Martinique diverged around 1988 (95% HPD, 1972 to 1992). In contrast, Dominican isolates were embedded within several clades of NM isolates (Fig. 3A). To further investigate the geographic spread of CC398 between the French overseas department Martinique and France, we carried out a phylogeographic reconstruction including French isolates from the Price et al. study (31) (Fig. 3B). This analysis revealed that the closest relatives of the Martinique strains are mostly from France, although not all nodes have high confidence in the predicted origin, which may reflect an undersampling of the strains of interest.

ST398 variable genome. MGEs are important contributors to virulence in *S. aureus* and to distinguishing livestock- and human-associated clades (24, 30, 31, 35). While LA-CC398 harbors an extensive repertoire of *S. aureus* phages, only a limited number of phages have been observed in human-CC398. Consistent with prior studies, mapping of prophages revealed that prophage SA3*int* was present in all human ST398 isolates, whereas prophages SA4*int* and SA8*int* were universally absent (see Fig. S3 in the supplemental material). Although only infrequently present in this sample collection, the Panton-Valentine leukocidin (PVL)-carrying phage SA2*int* was present in 12.5% (9/70) of infectious isolates but only 1.5% (2/139, $P = 0.0016$) of colonizing isolates (Fig. S3). In one household persistently colonized with two distinct ST398 clades, of which initially only one was carrying SA1*int*, we subsequently detected this phage in both clades, suggesting exchange of SA1*int* between isolates. Isolates from veterinarians and pigs in Minnesota, including those with the large-scale recombination, lacked SA3*int* and predominantly harbored SA6*int* ($n = 8$), consistent with an LA-ST398 origin.

We did not detect SNPs that uniquely defined the NM ST398 clade. However, a stop codon in *yeeR*, coding for a putative membrane protein, was uniquely associated with a major clade of the CC398 LA-MRSA isolates (Fig. S3).

Microevolution of ST398 within community households and networks. To investigate the transmission properties of ST398 in the community, we carried out a network
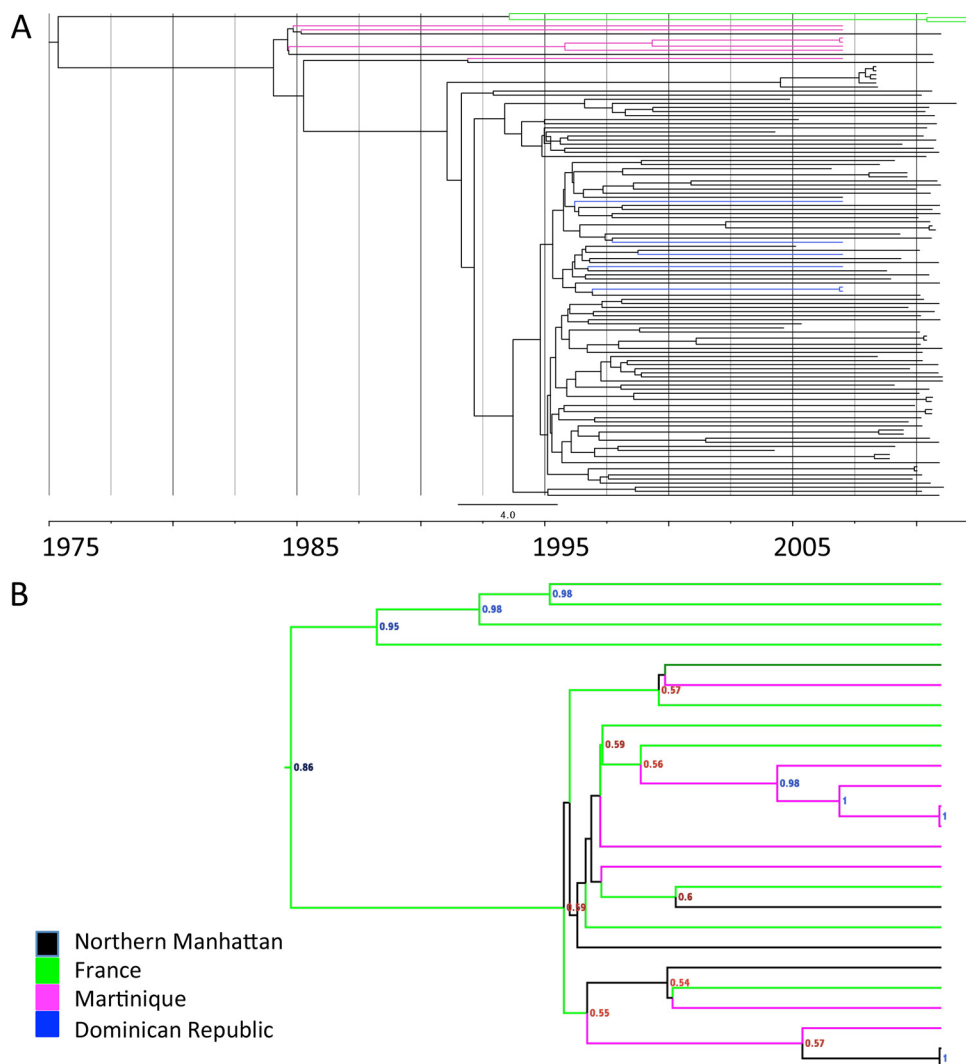
FIG 3 Bayesian phylogenetic reconstruction of ST398 isolates. (A) Maximum clade consensus tree, estimated from core genome SNPs of 119 isolates. Branches are scaled with time (months/years). Colors of internal and terminal branches indicate region of origin. (B) Phylogeographic spread of ST398 isolates between France and Martinique.

study of individuals previously identified as ST398 positive. We enrolled 15 networks that encompassed 348 participants from 81 different households. Thirty-three individuals from 20/81 (25%) households were colonized with t571-ST398 MSSA. ST398-colonized and noncolonized individuals were comparable in their demographics and comorbidities (see Table S2 in the supplemental material). They differed by only an increased household size and a decreased household density in the ST398 group as well as a higher frequency of recent SSTIs (16% versus 7%, P = 0.04). Fourteen participants (42%) remained longitudinally colonized with ST398 MSSA for the duration of their study participation of up to 12 months.

The majority of ST398 isolates collected from the same individual, household, or network clustered closely together in phylogenetic analyses as illustrated by short terminal branches (Fig. 2B). However, there were several exceptions in which select individuals (*n* = 2), households (*n* = 6), or networks (*n* = 5) harbored at least two highly distinct ST398 clades. Notably, in one large network with four ST398-colonized households, each harbored distinct ST398 subclades.

To further define the clonal diversity of individuals, households, and networks, we estimated the pairwise SNP distances between isolates. We found that the median pairwise SNP distance of isolates on the same individual was 8 (range, 0 to 80), that
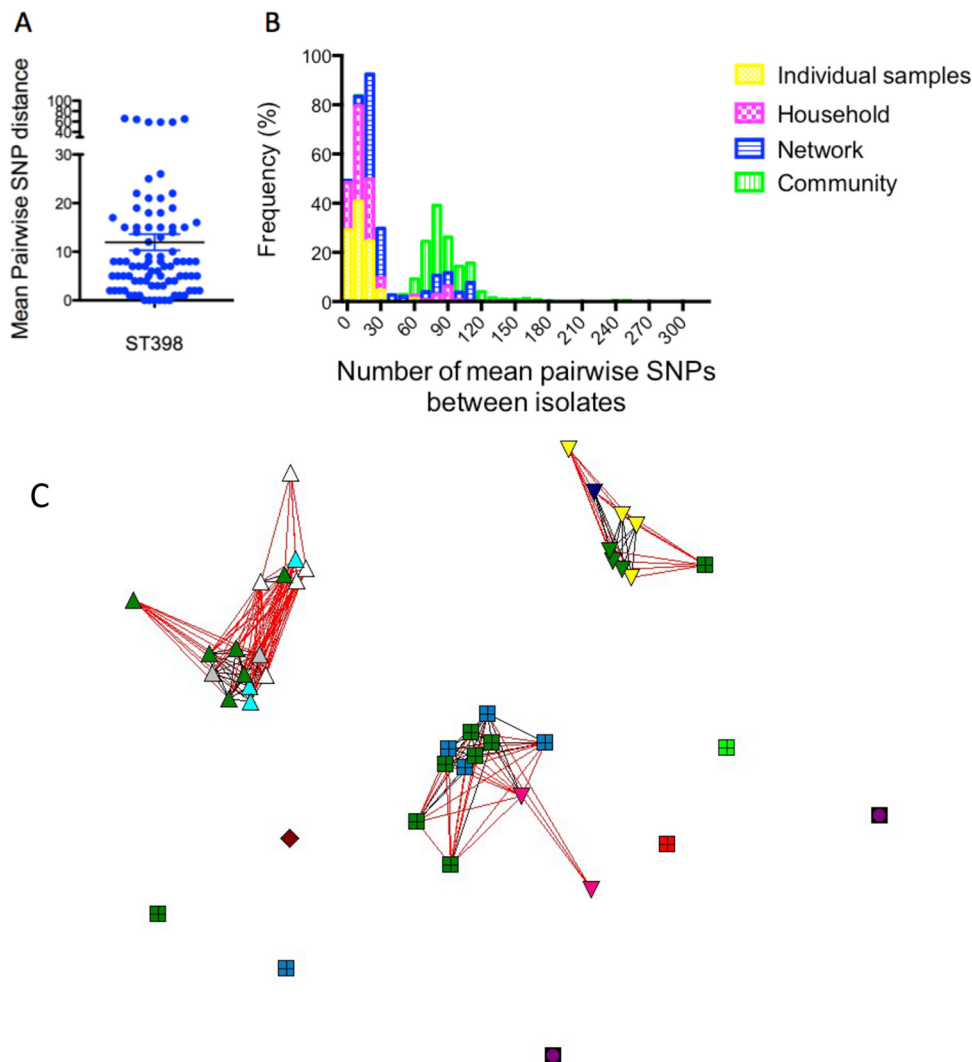
FIG 4 (A) Pairwise SNP comparisons between multiple samples taken from different body sites in individuals. (B) Mean pairwise SNPs between isolates from individuals (yellow), households (pink), or networks (blue) or between isolates from different community households (green). (C) SNP distance-based mapping of ST398 isolates from the largest social network indicates microevolution of separate clusters and frequent exchange between individuals and the environment and between households. Each symbol represents a single isolate, and isolates are clustered by their pairwise SNP distance. Each of the four embedded households is represented by a different shape; each individual is shown in a different color. All dark green symbols represent environmental samples. Isolates connected by red lines are separated by up to 10 pairwise SNPs. Nonconnected isolates are less than 25 SNPs different and are proportional to the number of SNPs.

within households was 13 (range, 0 to 105), that within networks was 25 (range, 3 to 114), and that in the community was 82 (range, 0 to 312) (Fig. 4).

We then applied these SNP diversity values to define and investigate transmission. While isolates mainly clustered by individual and household, there was extensive overlap of genotypes between individuals and households and across four networks, suggesting frequent exchange. Environmental isolates reflected most of the diversity present on individuals. Overall, these data indicate that closely interacting individuals share a cloud of isolate diversity that extends across households and networks (Fig. 2B and 4C).

These SNP-based diversity measures were significantly higher than those in our previous study on USA300-MRSA transmission in community households (individuals, 1.4 [range, 0 to 9], and households, 4 [range, 0 to 23]) (5), although substitution rates were comparable for the two clones (5, 32). We explored several possibilities for the relatively high SNP diversity. First, we compared longitudinal isolates that were col-

lected 6 months apart ($n = 38$). These follow-up samples did not have excess mutations based on the estimated substitution rate of 4 SNPs per year. We also did not observe mutations or premature stop codons in candidate genes of hypermutability such as *mutL/mutS*.

To explore deviations from neutral evolution within households and networks, we performed Tajima's *D* neutrality test (36). Following the method of Tajima (36), we assumed beta-distributed confidence limits for *D*. All but one household and network yielded negative *D* values (see Fig. S2). We identified significant nonrandom evolutionary processes in two networks: network 105 ($D = -1.95$, $P < 0.05$) and network 114 ($D = -2.25$, $P < 0.01$). Only one household in network 114 had a sufficient number of samples to recapitulate the network's significant results ($D = -2.01$, $P < 0.05$). Although we did not find values of *D* to be significant across genes, we found one pathway in network 105 and eight pathways in network 114 corroborating significant values of *D* from whole-genome sequences (see Table S3 in the supplemental material). We, however, found no evidence for purifying or positive selection, as the ratios of nonsynonymous to synonymous substitutions (*dN/dS*) were not significantly different from 1 across the study and within network-specific lineages. This suggests that negative *D* values are possibly due to the presence of rare alleles at low frequencies that accumulated during a recent selective sweep or an expansion after a bottleneck.

DISCUSSION

There is a lack of information on the evolution and spread of dominant MSSA lineages. This is one of the first longitudinal studies to investigate the evolution of the uniquely host-species-adaptable ST398 MSSA clone in the community using a combined social network and genomic approach. Similarly to recent household studies on USA300 MRSA (5, 37), we found that households served as major sites for ST398 MSSA transmission. However, by taking advantage of our network enrollment strategy, we were able to demonstrate frequent spread of these isolates within social networks as evident by overlapping SNP distances. We also found evidence for long- and short-distance geographic migration of ST398. First, isolates from infections that occurred during hospitalizations did not differ from colonizing isolates and were embedded within the clade of NM community isolates. This observation highlights the potential role of common community commensals in hospital-associated *S. aureus* infections. Second, isolates from France and the French overseas department Martinique were more closely related to evidence for spread from France to Martinique. French tourists account for the majority of visitors (75%) each year to Martinique (38). Bayesian phylogeographic reconstruction supported the root of Dominican ST398 isolates in Northern Manhattan and suggested subsequent spread to the Caribbean island, which is a frequent travel destination from this community (22). These links may help explain the disparate distribution of this clone in different geographic regions.

We observed a relatively high and overlapping SNP diversity within individuals, environmental surfaces, households, and networks. In our previous study on USA300 MRSA transmission in the same community, the SNP diversity was much lower on individuals and in households. However, the substitution rates observed here for ST398 as well as previously published rates for both clones (5, 32) are comparable. When analyzing longitudinal isolates collected from the same individual, we did not find any evidence for an increased accumulation of SNPs in ST398 isolates over this time period. We also did not detect SNPs in genes associated with hypermutability such as *mutL* or *mutS*.

Alternatively, differences in SNP diversity between lineages might be a direct reflection of long-term persistence of ST398-MSSA over years versus short-term colonization with USA300 in community households. This might be attributable to the increased virulence potential of USA300, the predominant CA-MRSA strain and major cause of SSTIs in the United States (39). Alam et al. (37), however, suggested long-term persistence of USA300 in households in Chicago and Los Angeles ranging between 2.3 and 8.4 years. In our study, we did not detect a temporal signal within households due

to the relatively large number of SNPs at each sampling time point, thus precluding timing of persistence. While both studies provide evidence for long-term persistence of successful *S. aureus* clones, it should be noted that these stand in contrast to rapid changes in individual colonization based on *spa* typing alone (40). The observed high SNP diversity in individuals has direct implications for epidemiological transmission studies, including in the health care setting, when patients become infected with their colonizing bacteria. Differences in SNP diversity between *S. aureus* clones such as ST398 and USA300 or other bacterial clones may need to be considered in defining linkages during hospital outbreaks.

Although we found no evidence for presence of selection from the *dN/dS* estimates, Tajima's *D* neutrality test suggests deviations from neutral evolution in households and networks. Negative Tajima *D* values indicated that the number of observed genetic differences (theta) is smaller than the expected value (pi), consistent with the presence of rare alleles at low frequencies, possibly due to a recent selective sweep or expansion after a bottleneck such as after transmission and adaptation to new colonization sites. Our study provides preliminary evidence that this might be driven by accumulation of SNPs in specific functional pathways. Despite differences in measuring genetic diversity between our study and that by Alam et al., both investigations observed that the number of observed genetic differences (theta) is smaller than the expected (pi), resulting in negative Tajima *D* values.

In a prior study on differences between human MSSA and LA-CC398 MRSA, we observed pseudogenes and variations in genes encoding surface adhesins in LA isolates but not in human samples (24). This further translated into impaired binding of LA-ST398 to human keratinocytes *in vitro*, whereas both human and LA isolates were able to adhere to porcine keratinocytes. Many of these mutations or insertions and deletions were detected in repetitive elements of these genes. Short-read sequences obtained by Illumina sequencing here preclude a thorough and reliable assessment of these regions.

We detected evidence for a large-scale recombination within a subset of ST398 MSSA isolates obtained from pigs and swine veterinarians in three Midwestern states. This is at least the second report of recombination in this clonal lineage (31). However, in contrast to large-scale recombinations previously described in ST398 and other *S. aureus* clonal complexes, the origin of this ~250-kb region cannot be attributed to a single clonal lineage. This indicates either recombination from a not-yet-identified novel sequence type or indeed mosaic acquisition of several regions from a diversity of clonal complexes (CCs). Homologous recombination might provide additional means of adapting to a novel host environment (34, 41).

Several limitations to our study need to be considered. Despite efforts to enroll complete social networks, we were able to recruit only about one-third of named close social contacts, introducing possible selection bias. Follow-up data were limited as individuals had moved or were not willing to participate any longer. Cost precluded a complete assessment of environmental samples. Phylogeographic reconstruction is vulnerable to sampling bias. More comprehensive sampling from other geographic origins may have changed the inferred origin of ST398.

In conclusion, we reconstructed the micro- and macroevolution of ST398-MSSA, a clone that is uniquely successful in diverse regions despite the lack of an array of antimicrobial resistance genes. Phylogeographic reconstruction suggests recent spread from France to Martinique and from NM to the DR, in accordance with cultural links and travel destinations (22), rather than spread within the Caribbean. We provide evidence for an extensive cloud of diversity of ST398-MSSA shared between individuals and the environment in households, extending to social networks, which was distinct from USA300. Recognizing these differences between clonal lineages has important implications for outbreak investigations, including in the clinical setting.

MATERIALS AND METHODS

Ethics statement. This study was reviewed and approved by the Institutional Review Board of Columbia University Medical Center. Written informed consent was obtained from each individual before

participation. Parental consent was required for the participation of children <18 years old. Pediatric assent was obtained from those capable of providing it.

Study population, demographics, and sample selection. As part of a community study on *S. aureus* transmission in Northern Manhattan, we implemented a cluster network analysis of ST398 transmission between September 2010 and March 2013. Individuals who had previously tested positive for ST398 colonization or infection were eligible to enroll (28, 42). These index participants were asked to complete a structured questionnaire and provide information on their daily contacts and the nature and frequency of these interactions. In addition, swabs were collected from the anterior nares, throat, and groin from consenting participants (Becton, Dickinson Culturette Systems). A standardized set of environmental items was also sampled with sterile premoistened swabs in all households: doorknobs, television remote control, living room light switch, toy, couch or bed, computer or radio, house phone or index cellular phone, bathroom sink, kitchen, and appliance handle (33).

A snowball enrollment strategy was used to identify contacts of index participants. Named household and outside-household contacts of the index were approached and after providing informed consent underwent the same study procedures. Contacts were further defined by their type of relationship (e.g., family, work, school, sports, or other social life contacts). Recruitment occurred in waves, such that contacts of study participants were asked to participate in the study, as were the contacts of these newly recruited contacts. Follow-up household visits were carried out twice: between 3 and 6 or 6 to 12 months after enrollment, respectively. We enrolled individuals from 15 networks that consisted of 958 named members. Of these, 348 were enrolled and 273 individuals agreed to complete the questionnaire. We compared demographic variables and risk factors for *S. aureus* infections between individuals colonized with ST398 and those colonized with other *S. aureus* isolates or not colonized. Data were analyzed using SAS 9.1 software (SAS Institute, Cary, NC).

Culture swabs were incubated overnight at 37°C in high-salt 6.5% broth and plated onto mannitol salt agar (Becton, Dickinson) for 48 h at 35°C. Positive mannitol-fermenting yellow colonies were isolated onto 5% sheep blood agar plates (Becton, Dickinson), and single colonies were selected for further analysis. *S. aureus* was identified by a coagulase and protein A detection kit (Murex Staphaurex). *S. aureus*-positive isolates were genotyped by *spa* sequencing using Ridom-staph software (24, 25). Of 2,590 swabs, we identified 690 *S. aureus* isolates. Based on *spa* typing and Spa clustering, 175 isolates belonged to Spa-CCt571, consistent with CC398.

Of these, we selected 153 isolates, including all human-colonizing isolates ($n = 96$) and a subset of environmental isolates (57/79) for sequencing. We also selected an additional 135 ST398 isolates from previous studies on *S. aureus* infection and colonization in Northern Manhattan and the Bronx as follows: (i) MSSA infections between 2007 and 2011 ($n = 69$) (24, 28), (ii) infectious isolates from the Dominican Republic and Martinique ($n = 14$) (22), (iii) community studies on colonization between 2003 and 2010 ($n = 38$) (23, 33, 42), and (iv) colonizing isolates from pigs in Minnesota and swine veterinarians in Minnesota, Illinois, and Indiana ($n = 14$) (26). In total, we obtained 288 novel sequences. Previously published sequences from international studies were also included in phylogenetic analyses (31, 32).

WGS and detection of SNPs in the core genome. *S. aureus* genomic DNA was extracted using the QIAamp DNA minikit (Qiagen), and unique index-tagged libraries were generated. WGS was carried out using the Illumina HiSeq2000 with 100-base paired-end reads. Paired-end reads were mapped against the core chromosome of the CC398 reference genome (GenBank accession number CP003045.1 [24]) using the Burrows-Wheeler Aligner (BWA) with the Smith-Waterman algorithm disabled as described previously (4). The same approach was used to assemble publicly available CC398 short reads. A core genome alignment was created from the consensus sequences, with the core genome defined as nucleotide sites shared by all sequences (alignment columns that did not contain a gap character or unknown nucleotide identity). Unmapped reads and sequences that were not present in all genomes and MGEs were not considered part of the core genome, and therefore, SNPs from these regions were not included in the phylogenetic analysis. For the ~250-kb region at the *ori*, we mapped sequence reads against known animal and human *S. aureus* reference genomes. Further regions that may be affected by recombination were tested using BratNextGen and removed. This ensured a curated high-quality data set for subsequent phylogenetic analysis comprised of 1,826,995 bp.

Phylogenetic analyses. Maximum-likelihood phylogenetic reconstruction of the CC398 isolate sequences was performed using RAxML v7.4.0 with the general time-reversible (GTR) model of nucleotide substitution and 1,000 bootstrap replicates (43). The resulting phylogenetic trees were rooted at the midpoint between the most divergent taxa in the tree. These analyses included either (i) the entire data set including previously published MSSA and MRSA genomes (376 isolates, 9,843 SNPs) or (ii) human ST398 MSSA sequenced in this study. Association of traits (host and location) with clades within the phylogeny was tested using the method described by Slatkin and Maddison (44). Trees were annotated using iTOL (45).

Time-scaled phylogenetic analysis. To investigate the temporal signal present in the maximum-likelihood phylogeny, the tree and dated tips were loaded into Path-O-Gen (http://tree.bio.ed.ac.uk/software/pathogen/) to identify isolates where the assumption of a molecular clock is not valid. A subset of 119 clinical isolates from the Manhattan clone displayed evidence of clocklike evolution and were used to infer the substitution rate, time of emergence, divergence dates, and phylogeographic spread, applying Bayesian methods implemented in the BEAST v1.7.5 package (46). BEAST was run for 100 million generations sampling every 10,000 states using the HKY substitution model; defined tip dates as in day, month, and year of isolation; and a strict, exponential relaxed, and lognormal relaxed molecular clock with constant size coalescent or Bayesian skyline coalescent, respectively. Each model was run three times, and good convergence of chains and effective sample size (ESS) values were inspected using

Tracer v1.5. A marginal likelihood estimation (MLE) using path sampling and steppingstone sampling for each run was carried out to compare the different combinations of clock and tree models (47, 48). The MLE was then used to assess the best-fitting model for the data set by calculating the Bayes factor. For the used data set, the strict skyline model provided the best fit. LogCombiner v1.7.5 was used to remove states before the burn-in and then to combine the trees from the multiple runs. A maximum clade credibility (MCC) tree from the combined trees was obtained using TreeAnnotator v1.7.5. To investigate transmission dynamics of *S. aureus* isolates from Martinique, discrete trait analysis was performed on the Martinique-associated clade, to determine the probability of the ancestral location at each node. BEAST was run for 100 million generations, sampling every 10,000 states, using the HKY substitution model, with the clock rate fixed at 1.

In silico detection of MGEs. Sequence reads were assembled *de novo* into contigs using Velvet v 1.0.12 (49) and Velvet Optimizer and were used to determine the presence and absence of MGEs. Prophages were classified according to their defined integrase type (50).

Transmission and molecular evolution analyses. We used pairwise single-nucleotide polymorphism (SNP) distances as a measure of genetic relatedness between isolates. Network relationships were visualized using Ucinet (51). We calculated Tajima's D test to compare the population mutation rate to the pairwise nucleotide distance within households and networks to determine whether the observed frequency of segregating mutations matched the expected frequency under the standard neutral model. We implemented calculation of Tajima's D in Matlab and assessed confidence intervals for D assuming the beta distribution as formalized by Tajima in 1989 (36). We calculated Tajima's D for households and networks for which at least 4 samples were available. We also calculated Tajima's D for genes and pathways based on the annotation of each genomic position using PATRIC's database for *Staphylococcus aureus* subsp. *aureus* 71193 (GenBank accession no. CP003045.1). We obtained estimates for ω, the dN/dS ratio, for the community study isolates as well as each network, based on a branch-site model (52) as implemented in the PAML package (53). For this analysis, we concatenated 778 protein-coding genes, which were found mutated in at least two isolates. To assess the presence of selection, we computed the likelihood-ratio test statistics for the estimated ω values versus fixed $\omega = 1$, using a chi-square test (54).

Accession number(s). Sequence data are deposited in the European Nucleotide Archive (study accession number PRJEB12818).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mBio.01375-16.

FIG S1, TIF file, 1.5 MB.
FIG S2, PDF file, 0.01 MB.
FIG S3, TIF file, 1.5 MB.
FIG S4, PDF file, 0.03 MB.
TABLE S1, XLSX file, 0.05 MB.
TABLE S2, XLSX file, 0.03 MB.
TABLE S3, XLSX file, 0.1 MB.

ACKNOWLEDGMENTS

REFERENCES

1. Lowy FD. 1998. Staphylococcus aureus infections. N Engl J Med 339: 520–532. https://doi.org/10.1056/NEJM199808203390806.
2. Harris SR, Feil EJ, Holden MT, Quail MA, Nickerson EK, Chantratita N, Gardete S, Tavares A, Day N, Lindsay JA, Edgeworth JD, de Lencastre H, Parkhill J, Peacock SJ, Bentley SD. 2010. Evolution of MRSA during hospital transmission and intercontinental spread. Science 327:469–474. https://doi.org/10.1126/science.1182395.
3. Holden MT, Hsu LY, Kurt K, Weinert LA, Mather AE, Harris SR, Strommenger B, Layer F, Witte W, de Lencastre H, Skov R, Westh H, Zemlicková H, Coombs G, Kearns AM, Hill RL, Edgeworth J, Gould I, Gant V, Cooke J, Edwards GF, McAdam PR, Templeton KE, McCann A, Zhou Z, Castillo-Ramírez S, Feil EJ, Hudson LO, Enright MC, Balloux F, Aanensen DM, Spratt BG, Fitzgerald JR, Parkhill J, Achtman M, Bentley SD, Nübel U. 2013. A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant Staphylococcus aureus pandemic. Genome Res 23:653–664. https://doi.org/10.1101/gr.147710.112.
4. McAdam PR, Templeton KE, Edwards GF, Holden MT, Feil EJ, Aanensen DM, Bargawi HJ, Spratt BG, Bentley SD, Parkhill J, Enright MC, Holmes A, Girvan EK, Godfrey PA, Feldgarden M, Kearns AM, Rambaut A, Robinson DA, Fitzgerald JR. 2012. Molecular tracing of the emergence, adaptation, and transmission of hospital-associated methicillin-resistant Staphylococcus aureus. Proc Natl Acad Sci U S A 109:9107–9112. https://doi.org/10.1073/pnas.1202869109.
5. Uhlemann AC, Dordel J, Knox JR, Raven KE, Parkhill J, Holden MT, Peacock SJ, Lowy FD. 2014. Molecular tracing of the emergence, diversification, and transmission of S. aureus sequence type 8 in a New York community. Proc Natl Acad Sci U S A 111:6738–6743. https://doi.org/10.1073/pnas.1401006111.

6. Stegger M, Wirth T, Andersen PS, Skov RL, De Grassi A, Simões PM, Tristan A, Petersen A, Aziz M, Kiil K, Cirković I, Udo EE, del Campo R, Vuopio-Varkila J, Ahmad N, Tokajian S, Peters G, Schaumburg F, Olsson-Liljequist B, Givskov M, Driebe EE, Vigh HE, Shittu A, Ramdani-Bougessa N, Rasigade JP, Price LB, Vandenesch F, Larsen AR, Laurent F. 2014. Origin and evolution of European Community-acquired methicillin-resistant *Staphylococcus aureus*. mBio 5:e01044-14. https://doi.org/10.1128/mBio.01044-14.

7. DeLeo FR, Kennedy AD, Chen L, Bubeck Wardenburg J, Kobayashi SD, Mathema B, Braughton KR, Whitney AR, Villaruz AE, Martens CA, Porcella SF, McGavin MJ, Otto M, Musser JM, Kreiswirth BN. 2011. Molecular differentiation of historic phage-type 80/81 and contemporary epidemic Staphylococcus aureus. Proc Natl Acad Sci U S A 108/81:18091–18096. https://doi.org/10.1073/pnas.1111084108.

8. Grundmann H, Aanensen DM, van den Wijngaard CC, Spratt BG, Harmsen D, Friedrich AW, European Staphylococcal Reference Laboratory Working Group. 2010. Geographic distribution of *Staphylococcus aureus* causing invasive infections in Europe: a molecular-epidemiological analysis. PLoS Med 7:e1000215. https://doi.org/10.1371/journal.pmed.1000215.

9. David MZ, Boyle-Vavra S, Zychowski DL, Daum RS. 2011. Methicillin-susceptible *Staphylococcus aureus* as a predominantly healthcare-associated pathogen: a possible reversal of roles? PLoS One 6:e18217. https://doi.org/10.1371/journal.pone.0018217.

10. Rolo J, Miragaia M, Turlej-Rogacka A, Empel J, Bouchami O, Faria NA, Tavares A, Hryniewicz W, Fluit AC, de Lencastre H, CONCORD Working Group. 2012. High genetic diversity among community-associated Staphylococcus aureus in Europe: results from a multicenter study. PLoS One 7:e34768. https://doi.org/10.1371/journal.pone.0034768.

11. Miko BA, Hafer CA, Lee CJ, Sullivan SB, Hackel MA, Johnson BM, Whittier S, Della-Latta P, Uhlemann AC, Lowy FD. 2013. Molecular characterization of methicillin-susceptible *Staphylococcus aureus* clinical isolates in the United States, 2004–2010. J Clin Microbiol 51:874–879 https://doi.org/10.1128/JCM.00923-12.

12. Sowash MG, Uhlemann AC. 2014. Community-associated methicillin-resistant *Staphylococcus aureus* case studies. Methods Mol Biol 1085:25–69. https://doi.org/10.1007/978-1-62703-664-1_2.

13. Armand-Lefevre L, Ruimy R, Andremont A. 2005. Clonal comparison of *Staphylococcus aureus* isolates from healthy pig farmers, human controls, and pigs. Emerg Infect Dis 11:711–714. https://doi.org/10.3201/eid1105.040866.

14. van Belkum A, Melles DC, Peeters JK, van Leeuwen WB, van Duijkeren E, Huijsdens XW, Spalburg E, de Neeling AJ, Verbrugh HA. 2008. Methicillin-resistant and -susceptible *Staphylococcus aureus* sequence type 398 in pigs and humans. Emerg Infect Dis 14:479–483. https://doi.org/10.3201/eid1403.070760.

15. Huijsdens XW, van Dijke BJ, Spalburg E, van Santen-Verheuvel MG, Heck ME, Pluister GN, Voss A, Wannet WJ, de Neeling AJ. 2006. Community-acquired MRSA and pig-farming. Ann Clin Microbiol Antimicrob 5:26. https://doi.org/10.1186/1476-0711-5-26.

16. van Loo I, Huijsdens X, Tiemersma E, de Neeling A, van de Sande-Bruinsma N, Beaujean D, Voss A, Kluytmans J. 2007. Emergence of methicillin-resistant *Staphylococcus aureus* of animal origin in humans. Emerg Infect Dis 13:1834–1839. https://doi.org/10.3201/eid1312.070384.

17. Cuny C, Nathaus R, Layer F, Strommenger B, Altmann D, Witte W. 2009. Nasal colonization of humans with methicillin-resistant *Staphylococcus aureus* (MRSA) CC398 with and without exposure to pigs. PLoS One 4:e6800. https://doi.org/10.1371/journal.pone.0006800.

18. Deiters C, Günnewig V, Friedrich AW, Mellmann A, Köck R. 2015. Are cases of methicillin-resistant *Staphylococcus aureus* clonal complex (CC) 398 among humans still livestock-associated? Int J Med Microbiol 305:110–113. https://doi.org/10.1016/j.ijmm.2014.11.007.

19. Larsen J, Petersen A, Sørum M, Stegger M, van Alphen L, Valentiner-Branth P, Knudsen LK, Larsen LS, Feingold B, Price LB, Andersen PS, Larsen AR, Skov RL. 2015. Meticillin-resistant *Staphylococcus aureus* CC398 is an increasing cause of disease in people with no livestock contact in Denmark, 1999 to 2011. Euro Surveill 20:30021. https://doi.org/10.2807/1560-7917.ES.2015.20.37.30021.

20. Rasigade JP, Laurent F, Hubert P, Vandenesch F, Etienne J. 2010. Lethal necrotizing pneumonia caused by an ST398 *Staphylococcus aureus* strain. Emerg Infect Dis 16:1330. https://doi.org/10.3201/eid1608.100317.

21. Vandendriessche S, Kadlec K, Schwarz S, Denis O. 2011. Methicillin-susceptible *Staphylococcus aureus* ST398-t571 harbouring the macrolide-lincosamide-streptogramin B resistance gene erm(T) in Bel-gian hospitals. J Antimicrob Chemother 66:2455–2459. https://doi.org/10.1093/jac/dkr348.

22. Uhlemann AC, Dumortier C, Hafer C, Taylor BS, Sánchez J, Rodriguez-Taveras C, Leon P, Rojas R, Olive C, Lowy FD. 2012. Molecular characterization of *Staphylococcus aureus* from outpatients in the Caribbean reveals the presence of pandemic clones. Eur J Clin Microbiol Infect Dis 31:505–511. https://doi.org/10.1007/s10096-011-1339-2.

23. Bhat M, Dumortier C, Taylor BS, Miller M, Vasquez G, Yunen J, Brudney K, Sánchez-E J, Rodriguez-Taveras C, Rojas R, Leon P, Lowy FD. 2009. *Staphylococcus aureus* ST398, New York City and Dominican Republic. Emerg Infect Dis 15:285–287. https://doi.org/10.3201/eid1502.080609.

24. Uhlemann AC, Porcella SF, Trivedi S, Sullivan SB, Hafer C, Kennedy AD, Barbian KD, McCarthy AJ, Street C, Hirschberg DL, Lipkin WI, Lindsay JA, DeLeo FR, Lowy FD. 2012. Identification of a highly transmissible animal-independent *Staphylococcus aureus* ST398 clone with distinct genomic and cell adhesion properties. mBio 3:e00027-12. https://doi.org/10.1128/mBio.00027-12.

25. Yu F, Chen Z, Liu C, Zhang X, Lin X, Chi S, Zhou T, Chen Z, Chen X. 2008. Prevalence of *Staphylococcus aureus* carrying Panton-Valentine leukocidin genes among isolates from hospitalised patients in China. Clin Microbiol Infect 14:381–384. https://doi.org/10.1111/j.1469-0691.2007.01927.x.

26. Sun J, Yang M, Sreevatsan S, Davies PR. 2015. Prevalence and characterization of *Staphylococcus aureus* in growing pigs in the USA. PLoS One 10:e0143670. https://doi.org/10.1371/journal.pone.0143670.

27. Brunel AS, Bañuls AL, Marchandin H, Bouzinbi N, Morquin D, Jumas-Bilak E, Corne P. 2014. Methicillin-sensitive *Staphylococcus aureus* CC398 in intensive care unit, France. Emerg Infect Dis 20:1511–1515. https://doi.org/10.3201/eid2009.130225.

28. Uhlemann AC, Hafer C, Miko BA, Sowash MG, Sullivan SB, Shu Q, Lowy FD. 2013. Emergence of sequence type 398 as a community- and healthcare-associated methicillin-susceptible *Staphylococcus aureus* in northern Manhattan. Clin Infect Dis 57:700–703. https://doi.org/10.1093/cid/cit375.

29. David MZ, Siegel J, Lowy FD, Zychowski D, Taylor A, Lee CJ, Boyle-Vavra S, Daum RS. 2013. Asymptomatic carriage of sequence type 398, spa type t571 methicillin-susceptible *Staphylococcus aureus* in an urban jail: a newly emerging, transmissible pathogenic strain. J Clin Microbiol 51:2443–2447. https://doi.org/10.1128/JCM.01057-13.

30. McCarthy AJ, Witney AA, Gould KA, Moodley A, Guardabassi L, Voss A, Denis O, Broens EM, Hinds J, Lindsay JA. 2011. The distribution of mobile genetic elements (MGEs) in MRSA CC398 is associated with both host and country. Genome Biol Evol 3:1164–1174. https://doi.org/10.1093/gbe/evr092.

31. Price LB, Stegger M, Hasman H, Aziz M, Larsen J, Andersen PS, Pearson T, Waters AE, Foster JT, Schupp J, Gillece J, Driebe E, Liu CM, Springer B, Zdovc I, Battisti A, Franco A, Zmudzki J, Schwarz S, Butaye P, Jouy E, Pomba C, Porrero MC, Ruimy R, Smith TC, Robinson DA, Weese JS, Arriola CS, Yu F, Laurent F, Keim P, Skov R, Aarestrup FM. 2012. *Staphylococcus aureus* CC398: host adaptation and emergence of methicillin resistance in livestock. mBio 3:e00305-11. https://doi.org/10.1128/mBio.00305-11.

32. Ward MJ, Gibbons CL, McAdam PR, van Bunnik BA, Girvan EK, Edwards GF, Fitzgerald JR, Woolhouse ME. 2014. Time-scaled evolutionary analysis of the transmission and antibiotic resistance dynamics of *Staphylococcus aureus* CC398. Appl Environ Microbiol 80:7275–7282 https://doi.org/10.1128/AEM.01777-14.

33. Uhlemann AC, Knox J, Miller M, Hafer C, Vasquez G, Ryan M, Vavagiakis P, Shi Q, Lowy FD. 2011. The environment as an unrecognized reservoir for community-associated methicillin resistant *Staphylococcus aureus* USA300: a case-control study. PLoS One 6:e22407. https://doi.org/10.1371/journal.pone.0022407.

34. Spoor LE, Richardson E, Richards AC, Wilson GJ, Mendonca C, Gupta RK, McAdam PR, Nutbeam-Tuffs S, Black NS, O'Gara JP, Lee CY, Corander J, Fitzgerald JR. 2015. Recombination-mediated remodelling of host-pathogen interactions during *Staphylococcus aureus* niche adaptation. Microb Genom https://doi.org/10.1099/mgen.0.000036.

35. Schijffelen MJ, Boel CH, van Strijp JA, Fluit AC. 2010. Whole genome analysis of a livestock-associated methicillin-resistant *Staphylococcus aureus* ST398 isolate from a case of human endocarditis. BMC Genomics 11:376. https://doi.org/10.1186/1471-2164-11-376.

36. Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595.

37. Alam MT, Read TD, Petit RA, III, Boyle-Vavra S, Miller LG, Eells SJ, Daum RS, David MZ. 2015. Transmission and microevolution of USA300 MRSA

in U.S. households: evidence from whole-genome sequencing. mBio 6:e00054-15. https://doi.org/10.1128/mBio.00054-15.

38. Caribbean Tourism Organization. 2010. 2003 - July 2015 tourism stats. Caribbean Tourism Organization, St. Michael, Barbados. http://www.onecaribbean.org/statistics/2003-july-2015-tourism-stats/.

39. Moran GJ, Krishnadasan A, Gorwitz RJ, Fosheim GE, McDougal LK, Carey RB, Talan DA, EMERGEncy ID Net Study Group. 2006. Methicillin-resistant *S. aureus i*nfections among patients in the emergency department. N Engl J Med 355:666–674. https://doi.org/10.1056/NEJMoa055356.

40. Miller RR, Walker AS, Godwin H, Fung R, Votintseva A, Bowden R, Mant D, Peto TE, Crook DW, Knox K. 2014. Dynamics of acquisition and loss of carriage of *Staphylococcus aureus* strains in the community: the effect of clonal complex. J Infect 68:426–439. https://doi.org/10.1016/j.jinf.2013.12.013.

41. Vos M. 2009. Why do bacteria engage in homologous recombination? Trends Microbiol 17:226–232. https://doi.org/10.1016/j.tim.2009.03.001.

42. Knox J, Uhlemann AC, Miller M, Hafer C, Vasquez G, Vavagiakis P, Shi Q, Lowy FD. 2012. Environmental contamination as a risk factor for intrahousehold *Staphylococcus aureus* transmission. PLoS One 7:e49900. https://doi.org/10.1371/journal.pone.0049900.

43. Stamatakis A, Ludwig T, Meier H. 2005. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics 21:456–463. https://doi.org/10.1093/bioinformatics/bti191.

44. Slatkin M, Maddison WP. 1989. A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics 123:603–613.

45. Letunic I, Bork P. 2011. Interactive Tree of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res 39:W475–W478. https://doi.org/10.1093/nar/gkr201.

46. Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 29:1969–1973. https://doi.org/10.1093/molbev/mss075.

47. Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. Mol Biol Evol 29:2157–2167. https://doi.org/10.1093/molbev/mss084.

48. Baele G, Li WL, Drummond AJ, Suchard MA, Lemey P. 2013. Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. Mol Biol Evol 30:239–243. https://doi.org/10.1093/molbev/mss243.

49. Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res 18:821–829. https://doi.org/10.1101/gr.074492.107.

50. Goerke C, Pantucek R, Holtfreter S, Schulte B, Zink M, Grumann D, Bröker BM, Doskar J, Wolz C. 2009. Diversity of prophages in dominant *Staphylococcus aureus* clonal lineages. J Bacteriol 191:3462–3468. https://doi.org/10.1128/JB.01804-08.

51. Borgatti SP, Everett MG, Freeman LC. 2002. Ucinet for windows: software for social network analysis. Analytic Technologies, Harvard, MA.

52. Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol 22:2472–2479. https://doi.org/10.1093/molbev/msi237.

53. Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591. https://doi.org/10.1093/molbev/msm088.

54. Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol 19:908–917. https://doi.org/10.1093/oxfordjournals.molbev.a004148.