Demography and Intercontinental Spread of the USA300 Community-Acquired Methicillin-Resistant *Staphylococcus aureus* Lineage

Philippe Glaser, Patricia Martins-Simões, Adrien Villain, Maxime Barbier, Anne Tristan, Christiane Bouchier, Laurence Ma, Michele Bes, Frederic Laurent, Didier Guillomet, Thierry Wirth, François Vandenesch

Institut Pasteur, Unité de Biologie des Bactéries Pathogènes à Gram-Positif, Paris, France; CNRS UMR3525, Paris, France; CIRI, International Center for Infectiology Research, Lyon, France; Inserm U1111, Lyon, France; Université Lyon 1, Lyon, France; École Normale Supérieure de Lyon, Lyon, France; CNRS UMR308, Lyon, France; CNR des Staphylocoques, Hospices Civils de Lyon, CBPE, Lyon, France; Institut Pasteur, Bioinformatics Platform, Paris, France; Laboratoire Biologie Intégrative des Populations, École Normale Supérieure de Lyon, Lyon, France; CNR des Staphylocoques, Hospices Civils de Lyon, Lyon, France; Université Lyon 1, Lyon, France; Institut Pasteur, Genomics Platform, Paris, France; École Pratique des Hautes Études, Sorbonne Universités, Paris, France; Institut Pasteur, Demographic and Intercontinental Spread of the USA300 Community-Acquired Methicillin-Resistant *Staphylococcus aureus* Lineage.

**ABSTRACT** Community-acquired methicillin-resistant *Staphylococcus aureus* (CA-MRSA) was recognized worldwide during the 1990s; in less than a decade, several genetically distinct CA-MRSA lineages carrying Panton-Valentine leukocidin genes have emerged on every continent. Most notably, in the United States, the sequence type 18-IV (ST8-IV) clone known as USA300 has become highly prevalent, outcompeting methicillin-susceptible *S. aureus* (MSSA) and other MRSA strains in both community and hospital settings. CA-MRSA bacteria are much less prevalent in Europe, where the European ST80-IV European CA-MRSA clone, USA300 CA-MRSA strains, and other lineages, such as ST22-IV, coexist. The question that arises is whether the USA300 CA-MRSA present in Europe (i) was imported once or on very few occasions, followed by a broad geographic spread, anticipating an increased prevalence in the future, or (ii) derived from multiple importations with limited spreading success. In the present study, we applied whole-genome sequencing to a collection of French USA300 CA-MRSA strains responsible for sporadic cases and micro-outbreaks over the past decade and United States ST8 MSSA and MRSA isolates. Genome-wide phylogenetic analysis demonstrated that the population structure of the French isolates is the product of multiple introductions dating back to the onset of the USA300 CA-MRSA clone in North America. Coalescent-based demography of the USA300 lineage shows that a strong expansion occurred during the 1990s concomitant with the acquisition of the arginine catabolic mobile element and antibiotic resistance, followed by a sharp decline initiated around 2008, reminiscent of the rise-and-fall pattern previously observed in the ST80 lineage. A future expansion of the USA300 lineage in Europe is therefore very unlikely.

**IMPORTANCE** To trace the origin, evolution, and dissemination pattern of the USA300 CA-MRSA clone in France, we sequenced a collection of strains of this lineage from cases reported in France in the last decade and compared them with 431 ST8 strains from the United States. We determined that the French CA-MRSA USA300 sporadic and micro-outbreak isolates resulted from multiple independent introductions of the USA300 North American lineage. At a global level, in the transition from an MSSA biotype to a successful CA-MRSA clone, it first became resistant to multiple antibiotics and acquired the arginine catabolic mobile element and subsequently acquired resistance to fluoroquinolones, and these two steps were associated with a dramatic demographic expansion. This expansion was followed by the current stabilization and expected decline of this lineage. These findings highlight the significance of horizontal gene acquisitions and point mutations in the success of such disseminated clones and illustrate their cyclic and sporadic life cycle.

*Staphylococcus aureus* remains one of the most challenging and costly sources of bacterial infection worldwide. It asymptomatically colonizes about one-third of the human population and may cause infections with outcomes that range from mild to severe and are occasionally life threatening (1). Notably, it is the most common causative agent of nosocomial infections and a leading cause of death in hospitalized patients (2). Until the mid-1990s, methicillin-resistant *S. aureus* (MRSA) infections were reported exclusively in hospital settings and most hospital-associated MRSA (HA-MRSA) diseases resulted from a limited number of successful clones (3). However, in the beginning of the 2000s, MRSA infections began to be reported in healthy individuals without risk factors or discernible connections to health care institutions (4, 5). These community-acquired MRSA (CA-
MRSA) strains had genetic backgrounds distinct from those of traditional HA-MRSA strains. Moreover, international strains of CA-MRSA belonged to a series of different lineages and specific clones were predominant on different continents (6). The dominant CA-MRSA clone in the United States, referred to as the USA300 North American (USA300-NA) MRSA clone in this study, belongs to pandemic multilocus sequence type 8 (ST8) constituted by methicillin-susceptible *S. aureus* (MSSA) and MRSA strains with the same pulsed-field gel electrophoresis type, USA300. This epidemiologically successful clone carries the IVa subtype of the staphylococcal cassette chromosome mec (SCCmec) element, agr allele 1, arginine catabolic mobile element (ACME) type I, and a set of virulence genes including lukSF-PV/lukF-PV, seq, and seq (7–10).

It is generally recognized that this USA300-NA MRSA clone descended from an ancestral USA500-like MSSA strain by the acquisition of various mobile genetic elements (MGEs) and shows very recent clonal expansion (2, 11). These MGEs, namely, SCCmec type IV carrying the mecA gene, *S. aureus* pathogenicity island 5 containing seq and seq, phage phiSA2 carrying the Panton-Valentine leukocidin genes, and ACME type I, are thought to contribute to the success and high virulence of the commonly known USA300-NA MRSA strains (12–16). ACME type I is found exclusively in the USA300-NA MRSA lineage (16–21). It has been shown that the functional modularity (arginine deiminase system *arc* and the speG gene, which encodes a polypeptide resistance enzyme) plays a major role in the enhanced success of this clone during colonization and skin infections (22, 23).

In recent years, multiple studies have shown that the USA300-NA MRSA clone has spread worldwide, with reports of outbreaks in South America, the Middle East, the western Pacific, and Europe (for a detailed review, see reference 24). Interestingly, in South America, the most prevalent CA-MRSA is a USA300 variant, the so-called Latin American variant or USA300-LV (i.e., ST8, spa type 008, SCCmec type IVc, lukSF-PV+, *arcA*) present in Colombia, Venezuela, and Ecuador (24–26).

It has recently been proposed that the USA300 MRSA lineage has evolved since the 1980s into two parallel epidemics, one in North America with the acquisition of the ACME by the ancestral USA300 lineage, forming the USA300-NA MRSA epidemic, and one in South America, with the acquisition of a novel copper and mercury resistance (COMER) element by the ancestral lineage (27).

In Europe, CA-MRSA epidemiology is characterized by clonal heterogeneity, although the most common European strain is the European clone (ST80-IV, *pvl+*) (28). This clone originally emerged from a West African ancestor (29) and expanded in the Mediterranean area, the Middle East, and North Africa, as many of the first patients infected with this clone in Europe had histories of travel to these regions. Although the prevalence of this “European” clone is relatively high in some countries, like Algeria and Tunisia, with a widespread distribution in both hospital and community settings (28–30), in Europe, the prevalence of this clone, as well as that of CA-MRSA in general, remains low (31, 32) and even seems to be declining (33, 34). This decay is in keeping with the calculation of a recent slow decrease in its effective population size based on a Bayesian skyline model (29). However, the question arose as to whether the CA-MRSA clone (USA300-NA or USA300-LV) could replace ST80 CA-MRSA and expand in a worrisome scenario similar to the one observed in the United States in the past decade. Consistent with this hypothesis is the fact that the USA300-NA MRSA clone has been reported to be present in many European countries for several years (31, 32, 35–45) and is occasionally associated with infection clusters (46–48). A recent study showed that an increase in CA-MRSA USA300 prevalence in 2013 in Geneva, Switzerland, was due to distinct importations from the North and South American continents (49).

This study addressed the question of the USA300 lineage’s capacity to expand successfully in Europe by applying whole-genome sequencing to a collection of French USA300 CA-MRSA strains responsible for sporadic cases, as well as micro-outbreaks, over the past decade, followed by a comparison with available genomes of USA300 CA-MRSA, USA300 MSSA strains, and other ST8 isolates from the United States (50, 51). Genome-wide phylogenetic relationships and coalescent-based analyses showed that the population structure of the French isolates reflects multiple introductions of the USA300-NA MRSA clone. Furthermore, the coalescent-based demography of USA300-NA lineage confirmed its success in the late 1990s but also suggested a sharp decline initiated around 2008, reminiscent of the rise-and-fall pattern observed in the ST80 lineage.

**RESULTS**

**USA300 phylogeography and resistance makeup.** Prior to phylogenetic reconstruction, we checked if the data set provided some evidence of homologous recombination. We first made a visual inspection of the concatenated single-nucleotide polymorphisms (SNPs), and no contiguous SNPs consisting of three or more SNPs, mirroring homologous recombination, were detected. In a second step, neighbor nets were inferred in order to detect putative recombination signatures that would result in networks rather than trees. No major splits were found, and the pairwise homoplasy index (PHI) test failed to detect recombination signatures (*P* = 0.475). We then performed whole-genome phylogenetic analysis of the 498 ST8 isolates (including MRSA, MSSA, and USA300-LV and USA300-NA MRSA isolates) obtained from France (*n* = 67) and the United States (*n* = 431) within a time span of 15 years (see Table S1 in the supplemental material). A total of 12,840 SNPs were used to generate a maximum-likelihood (ML) tree that shows, as previously reported, that strains from San Diego (51) and New York City (50) are interspersed (Fig. 1). The closest relative sister group of the USA300-NA lineage corresponded to seven strains, all isolated from the New York City area. These seven isolates carry the COMER element characteristic of the USA300-LV South American clade (27) and belong to this lineage. Adding French USA300 CA-MRSA strains to the phylogenetic analysis did not unravel any new branching pattern or hidden sublineages and revealed that all of the French isolates belong to the USA300-NA lineage. Among the U.S. isolates, a fluoroquinolone-resistant lineage with the same mutations in *gyrA* (Ser84Leu) and *parC* (Ser80Tyr) has emerged within the USA300-NA MRSA clone and disseminated globally (Fig. 1). We observed a similar distribution in France, with 18 fluoroquinolone-susceptible strains, including 10 from an outbreak in Le Puy-en-Velay (a town in central France) (47) and 46 resistant isolates, including the 28 isolates from an outbreak in a long-term care facility (Paris 16th District) showing the same two mutations in the *gyrA* and *parC* loci. Plasmid analysis showed that all except three French isolates carry plasmids related to previously described plasmid p18805-p03 (52). This plasmid encodes multiple antibiotic and heavy metal resistance genes [aminoglycosides, *ant*(6)-Ia and
**FIG 1** Phylogenetic reconstruction of ST8 and its derived USA300 lineage. ML tree based on 498 genomes and a total of 12,840 concatenated SNPs, rooted by using distantly related MSSA strain ERS092996. Branch colors indicate the geographic origins of the strains: red for New York State; yellow for California; green for Minnesota, and blue for France. Outer circles represent ACME and antibiotic resistance profiles, whereas the inner circle represents the years of isolation of the strains. The stars correspond to 12 predicted independent intercontinental USA300-NA CA-MRSA transfers from North America to France. The different assignments in the ST8 “sublineages” are also indicated by colored areas within the circular tree. Fq, fluoroquinolone.

aph(3’)-III; beta-lactams, blaz; macrolides mph(C); macrolides-lincosamides-streptogramin B, msr(A); cadmium, cadD and cadX. These plasmids showed almost 100% nucleotide sequence identity indicative of a single origin but different regions of deletion. Similar plasmids are also present in the vast majority of the American isolates (50, 51) but absent from the strains of the USA300-LV lineage. It is therefore likely that an ancestor of this multiresistance plasmid was acquired by the common ancestor of the USA300-NA clone and contributed to its expansion. The 28 isolates from the outbreak in a long-term care facility (Paris 16th District) carried in their chromosome a 3.2-kb transposable element containing the dfrr gene, which encodes a dihydrofolate reductase conferring trimethoprim resistance.

Temporal Bayesian analysis reveals the complexity of USA300 demography. In a second step, we used the Bayesian co-
a two-phase expansion event that perfectly matches the acquisitions clone. Interestingly, the USA300-NA clone went through a sharp bottleneck followed, giving rise to the so-called USA300-NA was relatively stable from the early 1960s to the mid-1990s. Then, the 498 ST8 genomes. The ST8 USA300 effective population size than that proposed by Planet et al. (around 1975) (27). The divergence with the USA300-LV MRSA South American lineage dates back to 1983, an estimate slightly more recent other. The divergence with the USA300-NA MRSA South American and antibiotic resistance that might have contributed directly to the success of the USA300-NA MRSA lineage.

FIG 2  Bayesian skyline plot indicating population size changes in the USA300 lineage over time with a relaxed molecular clock. The shaded area represents the 95% confidence interval, and the arrows point to major phenotypic events driven by ACME and antibiotic resistance that might have contributed directly to the success of the USA300-NA MRSA lineage.

We then generated a Bayesian skyline plot that estimated the phylogenetic time tree and to derive estimated dates for the common ancestors of the strains (see Fig. S1 in the supplemental material). The tree was calibrated by using the sampling dates of the isolates ranging from 2000 to 2014. We tested the performance of various demographic models that favored a Bayesian skyline model with a relaxed molecular clock (data not shown). The estimated short-term mutation rate corresponded to 1.34 × 10^{-6} (95% confidence interval of 1.11 × 10^{-6} to 1.56 × 10^{-6}) substitutions per nucleotide site per year, similar to that reported previously in USA300 MRSA (50, 54) and other S. aureus evolutionary lineages (55, 56). On the basis of this substitution rate, the estimated time of the most recent common ancestor (TMRCA) of the USA300-NA clone is 1996 (see Fig. S1 in the supplemental material), a result that slightly shifts the ancestor of the CA-MRSA clone toward the 21st century, compared to the estimates obtained by Uhlemann and colleagues (between 1970 and 1993) (50). In addition, our analyses revealed that the acquisition of ACME, methicillin resistance, and a multiresistance plasmid dates back to 1996 and that fluoroquinolone resistance was gained around 1998 (see Fig. S1 in the supplemental material). It is noteworthy that, because of the lack of intermediate strains harboring only one of those mobile elements, we could not determine whether the acquisition of the SCCmec element harboring the methicillin resistance gene mecA, the multiresistance plasmid, and the ACME type I element occurred simultaneously or one event occurred very shortly after the other. The divergence with the USA300-LV MRSA South American lineage dates back to 1983, an estimate slightly more recent than that proposed by Planet et al. (around 1975) (27).

We then calculated the genome-wide pairwise distances of the most closely related French-American strains and compared them with pairwise distances between strains collected from the same patient at multiple sites, multiple colonies from the same nasal sample, or strains collected from the same patient upon different hospital admissions (59, 60) (Fig. 4). The median number of SNP differences that separated the closest French-American epidemic link was 74 (range, 59 to 120) and significantly differed from the other two settings mentioned above (P > 0.0001). However, this number of differences remains smaller than the median number of SNP differences (n = 104) for USA300 CA-MRSA isolates collected from different households in a New York City community (50). In addition, bootstrap support for clades encompassing the French strains and their closest American homologs was >70% for >50% of the different subsamples (see Fig. S2 in the supplemental material). Therefore, the epidemic links we detected between American and French strains are likely to reflect direct, or at least very short, transmission paths.

The largest French cluster contains 28 isolates from an outbreak in a Parisian long-term care facility but also 3 strains isolated in Aubervilliers nearby Paris, as well as 6 unlinked isolates from different parts of France (Fig. 5A and B). This cluster is very likely the product of a single introduction from the United States. Its TMRCA is 1998 (Fig. 5C; see Fig. S1 in the supplemental material), suggesting that this subclone had been circulating for at least a decade in France before being noticed.
DISCUSSION

Here we show that, in most instances, the USA300 CA-MRSA cases identified in France in the last decade corresponded to sporadic and independent importations from the United States (USA300-NA MRSA clone) without further indication of spreading in French territory. This is consistent with another observation in Switzerland, where the local increased prevalence of USA300 CA-MRSA in 2013 in Geneva was shown to result from multiple importations from America (49). As the Institut de Veille Sanitaire (a disease surveillance agency of the Ministry of Health) strongly recommend referring strains associated with multiple cases of skin and soft tissue infections (SSTIs) (within a household or other communities) to the French National Reference Laboratory, it is thus likely that the lack of observed spreading truly is due to the absence or a very limited number of secondary cases and not to underreporting. The reasons for the apparent lack of success of these USA300-NA CA-MRSA isolates in diffusing within the French community are unknown. However, it is in keeping with the low prevalence of CA-MRSA observed in Europe and specifically in France (32), despite the isolation of various CA-MRSA lineages on the European continent as early as 1993 (38, 61). Moreover, these observations also argue against the hypothesis that the USA300-NA MRSA clone is intrinsically more successful than European CA-MRSA ST80 (3, 62, 63) and thus that the USA300-NA MRSA clone could be a potential threat in Europe.

In one instance of the present study, however, one imported lineage appeared to have a remarkable geographic spread, which was subsequently responsible for two outbreaks, one in a long-term care facility (Paris 16th District) and another limited outbreak in a facility for the disabled also in the Paris area (Aubervilliers) in 2013 and 2014 (Fig. 5A). The same lineage was also responsible for sporadic cases scattered in French territory, i.e., in Paris (75019, June 2011), northeastern France (Strasbourg, January 2013), the French Alps (Grenoble, September 2011), central France (Orleans, January 2011), and western France (St-Nazaire, June 2012). Our Bayesian analysis predicts a TMRCA for this clone of 1998 (Fig. 5C). This observation suggests that introduction of the USA300-NA MRSA clone can be successful but to a limited extent. We did not detect genes specific to this lineage or determine which genetic mechanisms might be behind the successful dissemination of certain isolates.

A recent report by Planet et al. (27) suggested that the history of imports from America (49). As the Institut de Veille Sanitaire (a disease surveillance agency of the Ministry of Health) strongly recommend referring strains associated with multiple cases of skin and soft tissue infections (SSTIs) (within a household or other communities) to the French National Reference Laboratory, it is thus likely that the lack of observed spreading truly is due to the absence or a very limited number of secondary cases and not to underreporting. The reasons for the apparent lack of success of these USA300-NA CA-MRSA isolates in diffusing within the French community are unknown. However, it is in keeping with the low prevalence of CA-MRSA observed in Europe and specifically in France (32), despite the isolation of various CA-MRSA lineages on the European continent as early as 1993 (38, 61). Moreover, these observations also argue against the hypothesis that the USA300-NA MRSA clone is intrinsically more successful than European CA-MRSA ST80 (3, 62, 63) and thus that the USA300-NA MRSA clone could be a potential threat in Europe.

In one instance of the present study, however, one imported lineage appeared to have a remarkable geographic spread, which was subsequently responsible for two outbreaks, one in a long-term care facility (Paris 16th District) and another limited outbreak in a facility for the disabled also in the Paris area (Aubervilliers) in 2013 and 2014 (Fig. 5A). The same lineage was also responsible for sporadic cases scattered in French territory, i.e., in Paris (75019, June 2011), northeastern France (Strasbourg, January 2013), the French Alps (Grenoble, September 2011), central France (Orleans, January 2011), and western France (St-Nazaire, June 2012). Our Bayesian analysis predicts a TMRCA for this clone of 1998 (Fig. 5C). This observation suggests that introduction of the USA300-NA MRSA clone can be successful but to a limited extent. We did not detect genes specific to this lineage or determine which genetic mechanisms might be behind the successful dissemination of certain isolates.

A recent report by Planet et al. (27) suggested that the history of
the USA300 CA-MRSA lineage followed a parallel evolution since the mid-1970s on the American continents; one epidemic sublineage disseminated in South America (so-called USA300-LV) and another disseminated in North America (referred to here as USA300-NA MRSA). The separation of these two sublineages is concomitant with the acquisition of the ACME by the North American USA300 MRSA clone, while USA300-LV acquired another mobile element—the COMER element. They also acquired, likely independently, two different subtypes of the SCCmec elements, type IVc in the LV clade and type IVa in the NA clade. The demography of the USA300-NA lineage based on Bayesian analysis is in agreement with this model, showing a considerable demographic expansion around 1996, which coincides with the acquisition of the ACME and the SCCmec element by this USA300-NA
lineage, followed 2 years later by a second expansion peak that likely corresponds to the acquisition of fluoroquinolone resistance. This period of expansion assessed by Bayesian analysis is in line with the observed global spread of the USA300-NA CA-MRSA clone in North America and worldwide, although with a more limited impact in Europe and the Paris 16th District be-

Intercontinental Phylogeography of USA300

CA-MRSA between July 2013 and February 2014. The other out-

break occurred in a long-term care facility in the Paris 16th District be-

The first outbreak occurred in Le Puy en Velay, a small town located in a rural area in Auvergne, France, and involved 12 individuals with community-acquired SSTIs, i.e., 6 children attending the same day care facility for the concomitant emergence and expansion of the various CA-MRSA clones in different countries or on different contin-

in the Paris area; four patients were detected as carriers or infected with community-acquired SSTIs, i.e., 6 children attending the same day care facility for the concomitant emergence and expansion of the various CA-MRSA clones in different countries or on different continents at the end of the 20th century (5).

The most plausible scenario explaining CA-MRSA clone dy-

namic evolution is mostly clonal and that recom-

bination can be neglected. Therefore, in a preliminary step, we tested for putative recombination signatures were inferred with Neighbor-Net (86). Putative recombination signatures were inferred with Neighbor-Net (86).

Most of the analyses developed in our analytical framework (phylogenetics and Bayesian inference) are based on the assumptions that S. aureus evolution is mostly clonal and that recom-

bination can be neglected. Therefore, in a preliminary step, we tested for the presence of mosaic genomes with the algorithm SplitsTree v4.13.1 (86). Putative recombination signatures were inferred with Neighbor-Net (87), and each data set was analyzed for the presence of recombinant sequences with the PHI test in SplitsTree with an alpha value of 0.001.

Phylogenetic analyses. Phylogenetic reconstructions were per-

formed by considering the 12,840 polymorphic sites retained in the core genome of the 498 isolates. Phylogenetic relationships were reconstructed by the ML approach implemented in PhyML 3.0 (88). The robustness of the ML tree topology was assessed with bootstrapping analyses of 1,000 pseu-

Published genomic data used for comparative analysis. Genomic data from 36 isolates from the first large-scale genomic studies on USA300 CA-MRSA were retrieved (51). These strains had been isolated between 2003 and 2007 in California. In addition, the genomic data from 387 ST8 isolates from an extensive analysis of a New York City community performed between 2009 and 2011 were included. Metadata from these two studies were graciously provided by the authors (50, 51).

DNA sequencing and SNP detection. S. aureus genomes were se-

quenced by using the Illumina Hiseq 2000 (101 nucleotide reads) or MiSeq (150 nucleotide reads) sequencer, with a coverage of >75×. Libraries were constructed with the Illumina TruSeq kit. Sequence reads from previously described USA300 strains were downloaded from the European Nucleotide Archive website (http://www.ebi.ac.uk/ena) or obtained directly from the authors. Sequence reads were aligned with the first completely sequenced and annotated USA300-NA genome, FPRJ757 (GenBank accession no. CP000255.1) (10) by using the Burrows-Wheeler Alignment tool (BWA mem 0.7.5a) (80). SNP calling was done with the Genome Analysis Toolkit (GATK 2.7-2) (81) Unified Genotyper by following Broad Institute best practices. Candidate SNPs were further filtered by requiring coverage of greater than half of the genome mean coverage and 95% read agreement to validate the call. SNPs selected with this stringent filter were regarded as true positives and were searched for in the original set for every strain in which they had been filtered out. SNPs, short indels, and coverage were visualized with SynTView (82). Specific analysis of clones from the main French lineage was done by manually checking read alignments around SNPs with Tablet (83). Four regions corresponding to MGEs not present or showing a high density of SNPs in other USA300 isolates and in closely related ST8 strains were removed from the analysis, leaving a total of 12,840 polymorphic sites. These regions are the ACME and the SCCmec cassette (41 to 125 kb), pathogenic-

ity island 3 (encoding enterotoxins K and Q; 881 to 896 kb) (84), and integrated phages phiSA2usa (encoding the Panton-Valentine toxin; 1,546 to 1,644 kb) and phiSAusa (2,084 to 2,127 kb). These regions represent a total of 240 kb or 8.3% of the strain USA300_FPRJ757 chromosome.

Accessory genome analysis. Plasmids and chromosomal regions specific to the French isolates compared to the FPRJ757 reference genome were analyzed following de novo assembly of unmapped reads. Assembly was performed with Velvet by using an optimized k value (85). Plasmids and phages were identified by BLASTn searches with plasmid and phage sequences from S. aureus as the queries. Antibiotic resistance genes were identified with the Center for Genomic Epidemiology web tool (http://www.genomicepidemiology.org/).

Recombination detection. Most of the analyses developed in our analytical framework (phylogenetics and Bayesian inference) are based on the assumptions that S. aureus evolution is mostly clonal and that recom-

bination can be neglected. Therefore, in a preliminary step, we tested for the presence of mosaic genomes with the algorithm SplitsTree v4.13.1 (86). Putative recombination signatures were inferred with Neighbor-Net (87), and each data set was analyzed for the presence of recombinant sequences with the PHI test in SplitsTree with an alpha value of 0.001.

Phylogenetic analyses. Phylogenetic reconstructions were per-

formed by considering the 12,840 polymorphic sites retained in the core genome of the 498 isolates. Phylogenetic relationships were reconstructed by the ML approach implemented in PhyML 3.0 (88). The robustness of the ML tree topology was assessed with bootstrapping analyses of 1,000 pseu-

MATERIALS AND METHODS

Bacterial strains. The bacterial strains used in this study were from 24 sporadic cases that were geographically and temporally dispersed and from three limited outbreaks reported to the French National Reference Center for Staphylococci (CNR Staphylococci). All strains were isolated from cases (infection and/or carriage) that occurred in continental French territory in the past 11 years (see Table S1 in the supplemental material). The first outbreak occurred in Le Puy en Velay, a small town located in a rural area in Auvergne, France, and involved 12 individuals with community-acquired SSTIs, i.e., 6 children attending the same day care facility and 6 adult relatives in six households, between June 2011 and May 2012 (47). Two outbreaks occurred in long-term health care units. One was in a nursing home for mentally disabled patients in Aubervilliers in the Paris area; four patients were detected as carriers or infected with USA300 CA-MRSA between July 2013 and February 2014. The other out-

break occurred in a long-term care facility in the Paris 16th District between June 2012 and December 2013 and involved 10 infected patients and 12 carriers.

Four strains from the United States corresponded to sporadic cases of SSTIs that occurred before April 2003 and were provided by Barry Kre-

iswirth (77); three other strains originated from patients in Minnesota before 2001 and were provided by Timothy Naimi (78). The genomic backgrounds of these strains were assessed by diagnostic DNA microarray analysis (Identibac S. aureus Genotyping; Alere, Jena, Germany) as previously described (79). This microarray covers 332 different target sequences corresponding to 185 distinct genes and their allelic variants. The affiliation of isolates with clonal complexes was determined by a comparison of their hybridization profiles with those of reference strains previously characterized by multilocus sequence typing (79).
doreplicated datasets. A transversion substitution model was selected on the basis of Akaike’s information criterion with jModelTest 2.1.3 (89). Phylogenies were rooted with MSSA strain ERS092996.

Coalescent-based analyses. Evolutionary rates and tree topologies were analyzed with the generalized time reversible (GTR) and Hasegawa-Kishino-Yano (90) (HKY) substitution models with gamma distributed among-site rate variation with four rate categories (F4). We tested both a strict molecular clock (which assumes the same evolutionary rate for all of the branches of the tree) and a relaxed clock that allows different rates among the branches. Constant-size, logistic, exponentially growing coalescent models were used. We also considered the Bayesian skyline plot model (91), based on a general, nonparametric prior that enforces no particular demographic history. We used a piecewise linear skyline model with 20 groups and then compared the marginal likelihood of each model with Bayes factors estimated in Tracer 1.5. Bayes factors represent the ratio of the marginal likelihood of the models being compared. Approximate marginal likelihoods for each coalescent model were calculated via importance sampling (1,000 bootstrap replications) with the harmonic mean of the sampled likelihoods. A ratio between 3 and 10 indicates moderate support of the idea that one model fits the data better than another, whereas values of >10 indicate strong support. For each analysis, two independent runs of 100 million steps were performed and the chain was sampled every 10,000th generation. Examination of the Markov chain Monte Carlo (MCMC) samples with Tracer 1.5 indicated convergence and adequate mixing of the Markov chains, with effective sample sizes for each parameter in the hundreds or thousands. The first 10% of each chain was discarded as burn-in. We found the maximum clade credibility topology with TreeAnnotator 1.7.5 (52), and we reconstructed the Bayesian skyline plot with Tracer 1.5. The relaxed clock models provided a better fit to the data (Bayes factor, >12) and under the different models tested, the Bayesian skyline model provided a (marginally) better fit, overall.

Analysis of genes under positive selection. To capture SNPs under positive selection, we used the PCAdapt software to perform a genome scan based on a Bayesian factor model (57). We chose K = 2 factors because the third and the fourth factors did not correspond to population structure and distinguished individuals within the same clades. The factor analysis was performed on the centered genotype matrix that was not scaled. The MCMC algorithm was initialized by singular value decomposition, and the total number of steps was 400 with a burn-in of 200 steps.

SUPPLEMENTAL MATERIAL

Figure S1, PDF file, 0.9 MB.
Figure S2, PDF file, 0.1 MB.
Table S1, XLS file, 1.7 MB.
Table S2, DOCX file, 0.1 MB.

ACKNOWLEDGMENTS
We are grateful to all of the clinicians and microbiologists from the different laboratories who provided samples and/or clinical data included in this study. In particular, we gratefully acknowledge Tim Naimi and Barry Kreiswirth in the United States for providing isolates, Beate Heym and Grenellen Baud and Olivier Lesens of the Clermont-Ferrand Hospital for their contribution to the investigation of the two first major outbreaks that took place in France in a long-term care facility in the Paris area and in a day care center at Le Puy en Velay and for providing us with all of the isolates involved in these outbreaks. We thank the technicians of the National Reference Center for Staphylococci for their skillful technical contribution and Jerome Etienne and Gerard Lina for scientific input. We thank Pierre Lechat of the bioinformatics platform of the Pasteur Institute for SynTView integration. We also thank Anne-Catrin Uhlemann and Ryan Tewhey, who graciously provided metadata concerning the USA300 genomes used in their studies (30, 31).

This work was supported by the Fondation pour la Recherche Médicale (grant IGN20111223510); Labex ECOFECT, Labex IBEID, and Infrastructure d’Avenir France Génomique (ANR10-IBSN-09-08); and the Institut de Veille Sanitaire (INVS).


Glaser et al.


73. Landrum ML, Neumann C, Cook C, Chukwuma U, Ellis MW, Hos-


