



Molecular Epidemiology of Community-Associated Methicillin-resistant *Staphylococcus aureus* in the genomic era: a Cross-Sectional Study

Mattia Proserpi^{1,2}, Nazle Veras^{1,2}, Taj Azarian³, Mobeen Rathore⁴, David Nolan^{1,2}, Kenneth Rand¹, Robert L. Cook³, Judy Johnson^{1,2}, J. Glenn Morris Jr.² & Marco Salemi^{1,2}

¹College of Medicine, Department of Pathology, Immunology and Laboratory Medicine, University of Florida, ²Emerging Pathogens Institute, University of Florida, ³College of Public Health and Health Professions and College of Medicine, Department of Epidemiology, University of Florida, ⁴Division of Pediatric Infectious Diseases and Immunology, Department of Pediatrics, University of Florida College of Medicine-Jacksonville, Jacksonville, FL.

Methicillin-resistant *Staphylococcus aureus* (MRSA) is a leading cause of healthcare-associated infections and significant contributor to healthcare cost. Community-associated-MRSA (CA-MRSA) strains have now invaded healthcare settings. A convenience sample of 97 clinical MRSA isolates was obtained from seven hospitals during a one-week period in 2010. We employed a framework integrating *Staphylococcus* protein A typing and full-genome next-generation sequencing. Single nucleotide polymorphisms were analyzed using phylodynamics. Twenty-six t002, 48 t008, and 23 other strains were identified. Phylodynamic analysis of 30 t008 strains showed ongoing exponential growth of the effective population size the basic reproductive number (R_0) ranging from 1.24 to 1.34. No evidence of hospital clusters was identified. The lack of phylogeographic clustering suggests that community introduction is a major contributor to emergence of CA-MRSA strains within hospitals. Phylodynamic analysis provides a powerful framework to investigate MRSA transmission between the community and hospitals, an understanding of which is essential for control.

Staphylococcus aureus is a causative agent of skin and soft tissue infections (SSTI) and invasive disease with high rates of morbidity and mortality¹. *S. aureus* is also the leading cause of hospital-associated infections (HAI)²⁻⁴, contributing significantly to increased healthcare costs⁵. In 2008, the latest year of available data, CDC estimated that MRSA was responsible for 89,785 cases of invasive disease causing 15,249 deaths in the US⁶. The Center for Medicaid and Medicare Services no longer reimbursing excess hospitals charges attributed to HAIs compounds the financial impact of this issue⁷.

Over the past 70 years, since the discovery and widespread utilization of antibiotics, multi-drug resistant strains of *S. aureus* have emerged. Methicillin-resistant *S. aureus* (MRSA) originally appeared in hospitals in the 1960s, and then reemerged in the community and hospitals in the 1990s, spreading worldwide and creating reservoirs in both settings⁸. Until the mid-1990s, MRSA infections were mostly reported among individuals with predisposing risk factors and exposure to healthcare facilities (HCF)⁹. However, over the past fifteen years in the United States, we have witnessed a dramatic increase of community-associated (CA) cases in healthy people lacking known risk factors or exposure to the healthcare system¹⁰. CA-MRSA strains are genetically distinct compared to healthcare-associated MRSA (HA-MRSA) strains. Particularly, CA-MRSA isolates tend to be resistant to fewer non- β -lactam antibiotics, carry a smaller version of the genetic region responsible for methicillin resistance (SCC*mec* IV or SCC*mec* V), and often produce the Pantone-Valentine leukocidin (PVL)^{11,12}. In the United States, CA-MRSA strains also seem to spread more efficiently in community settings and are more virulent than HA-MRSA strains¹³⁻¹⁵.

It was previously thought that CA-MRSA strains were isolated to populations outside of the healthcare setting and caused relatively mild infections limited to uncomplicated SSTIs¹⁶. More recently, we have observed a blurring of the definitions which delineate CA- and HA- MRSA both molecularly and epidemiologically. MRSA strains with molecular characteristics of CA-MRSA have invaded healthcare settings and are now recognized as an important cause of HAIs¹⁷. In 2008, almost 27% of hospital-acquired MRSA infections were

SUBJECT AREAS:
PHYLOGENETICS
EPIDEMIOLOGY
BACTERIAL INFECTION
PATHOGENS

Received
23 November 2012

Accepted
7 May 2013

Published
28 May 2013

Correspondence and
requests for materials
should be addressed to
M.S. (salemi@
pathology.ufl.edu)



due to USA300 strains⁶. Within some healthcare institutions, CA-MRSA strains have replaced HA-MRSA strains¹⁰. These events demonstrate that current infection control measures have failed to prevent the emergence of CA-MRSA strains from becoming a major contributor to HAIs.

Increasing colonization pressure, or the proportion of patients infected or colonized with MRSA upon entry to a HCF, is identified as a major driving force for the emergence of CA-MRSA as a cause of HAIs^{10,17,18}. As the proportion of patients admitted to HCFs with MRSA increases, so does the opportunity for nosocomial transmission^{18,19}. This nosocomial transmission additionally exposes CA-MRSA strains, previously susceptible to a wider range of antibiotics than HA- strains, to greater selective antibiotic pressure. While colonization pressure contributes to the MRSA burden within HCFs, little is known about the changing dynamics of MRSA in the community and how these changes are affecting nosocomial transmission. Understanding the dynamics of MRSA at the interface of the hospital and in the community is critical to evaluating current prevention measures and designing effective interventions¹⁰.

Molecular characterization methods are an essential component in the study of pathogen epidemiology and allow discrimination between isolates of epidemiologically important organisms. A variety of molecular typing methods can be independently used to classify MRSA strains, including pulsed-field gel electrophoresis (PFGE), multilocus sequence typing (MLST), or *spa*-typing by sequencing the highly polymorphic *Staphylococcus* protein A (*spa*) gene²⁰. CA-MRSA strains in the United States are most commonly in a genetic cluster designated as PFGE type USA300, MLST type ST8, or *spa*-type t008^{21,22}. Additionally, Healthcare-Associated strains most commonly cluster in PFGE type USA100, also recognized as *spa*-type t002. *Spa* type distribution and frequencies have been used in a number of studies to characterize MRSA epidemics since they provide moderate discrimination and possess high throughput and good inter-laboratory reproducibility^{12,23–26}. *Spa* typing can characterize *S. aureus* isolates within a defined setting and identify potential epidemiological clusters by providing limited discrimination within a clonal complex. However, greater discrimination such as provided with whole-genome sequencing (WGS) and single nucleotide polymorphism (SNP) analysis would be useful, for example to discern outbreak from non-outbreak strains in settings where sporadic strains of a specific *spa*-types (i.e. t008) are common.

Molecular clock-calibrated phylogenetic trees make it possible to investigate the ancestral population from which a given pathogen originated and the evolutionary as well environmental factors contributing to successful epidemic spread²⁷. Such studies have mostly been limited to fast evolving viruses because they require genomic sequences, sampled over relatively short time intervals, displaying sufficient diversity in order to infer reliable phylogenies^{28–31}. However, the recent application of next generation full-genome sequencing and phylogenetic analysis to the study of bacterial pathogens has demonstrated the ability to discriminate between extremely similar organisms collected within a short timeframe^{32–34}. Additionally, high-resolution phylogenetic and phylogeographic (phylogenetic) analyses based on genome-wide SNP data are a powerful tool to infer the origin and test spatiotemporal hypotheses of MRSA spread^{32,33}. Such analyses can reveal temporally and spatially related isolates, elucidate the epidemiology of MRSA transmission in the community, and identify reservoirs when combined with epidemiological data. In addition, this facilitates an understanding of pathogen success in terms of emergence, virulence, and epidemics⁸. Currently, whole-genome analysis of *S. aureus* has largely been limited to *in vitro* drug resistance³⁵, bacterial population size³⁶, and geographic distribution of different strains³⁷.

This pilot study was designed to investigate whether MRSA circulation in northeast Florida hospitals is the result of hospital-specific epidemics (i.e. endemic transmission) or heterogeneous

mixing of strains potentially originating from community reservoirs. In order to test this hypothesis, we applied an innovative framework based on phylodynamic analysis integrating molecular *spa* typing and full-genome next-generation sequencing data by Illumina. In particular, we measured the degree of hospital-specific clustering of MRSA strains, as well as the bacterial gene flow (migration) among different hospitals. Several investigators have used WGS for the study of emerging pathogens^{8,38–40}. Although the analysis was based on a convenience sample and general conclusions should be interpreted accordingly, it clearly shows how sophisticated molecular epidemiology tools, until now mainly used to track outbreaks of fast evolving viruses, can successfully be applied to analyze full genome data of emerging bacterial pathogens.

Results

We analyzed a convenience sample of 97 clinical MRSA isolates from six hospitals in northeast Florida. Overall, 48 (50%) were classified as *spa* type t008, 26 (27%) as t002, 23 (24%) as other types/unknown types (Table 1). Out of the 59 isolates from Jacksonville, 42 (71%) were t008, 4 (7%) were t002, and 13 (22%) were other/unknown types. While information regarding isolate source was not specifically requested, two of the Jacksonville facilities reported that 25 samples (42.4%) were from inpatients. The remaining 34 isolates could have originated from inpatients or outpatients. Among the 38 isolates from Gainesville, 6 were t008 (16%), 22 were t002 (68%), and were 10 (26%) other types/unknown. Figure 1 summarizes distribution of *spa* types across hospitals. Compared to Gainesville, t008 isolates were more prevalent in Jacksonville ($p < 0.0001$). Furthermore, among all sampled Jacksonville hospitals, the proportion of t008 isolates within each facility was not significantly different ($p = 0.156$) compared to other types.

After processing the next-generation sequencing data, a final multiple alignment of 40 t008 MRSA isolates (26 from Jacksonville, four from Gainesville and 10 from GenBank) including 3,249 SNPs was generated (Supplementary file 1). Eleven of the 26 Jacksonville isolates (42.3%) were confirmed to have come from inpatients. The remaining 15 isolates may have originated from inpatients or outpatients. A preliminary analysis of phylogenetic signal using a transition/transversion vs. divergence graph and the Xia's test ($p < 0.0001$) did not show evidence for substitution saturation. This indicated that enough signal for phylogenetic inference existed

Table 1 | Frequencies of *spa*-types across seven northeast Florida hospitals, collected during 2010 (n = 97)

<i>Spa</i> -type	Number	Percent
t002	26	26.8%
t007	1	1.0%
t008	49	50.5%
t010	1	1.0%
t024	1	1.0%
t045	4	4.1%
t064	1	1.0%
t067	1	1.0%
t078	1	1.0%
t105	2	2.1%
t1062	1	1.0%
t1107	1	1.0%
t121	1	1.0%
t2229	1	1.0%
t242	1	1.0%
t539	1	1.0%
t688	1	1.0%
t746	1	1.0%
Unknown	2	2.1%
Total:	97	100%

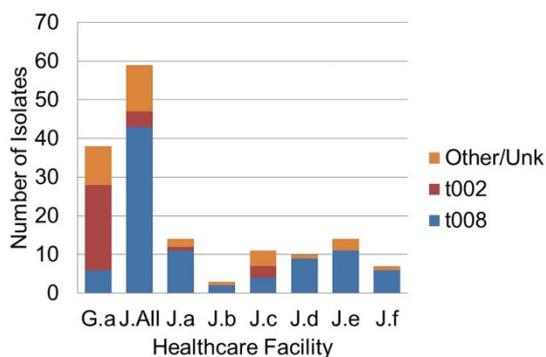


Figure 1 | Distribution of MRSA spa types across six different hospitals in Jacksonville (J) and one in Gainesville (G), both in northeast Florida, USA, collected during 2010 (n = 97).

(Supplementary Figure S1, and Supplementary Table T1). Likelihood mapping analysis reported < 25% of star-like signal (phylogenetic noise) and no significant signal for recombination was detected (PHI test $p = 0.82$, supplementary figure S2 and S3), indicating overall that

the data set contained enough information for reliable phylogeny inference.

The optimal evolutionary model as selected by the Akaike information criterion using MEGA5 was the general time reversible (GTR). Figure 2 depicts the GTR maximum-likelihood (ML) tree, including bacterial sequences from Jacksonville and Gainesville, shows no distinct clustering of hospital-specific clades. All trees estimated with other methods and including GenBank reference sequences are available as supplementary material and showed exactly the same pattern.

Strict and relaxed molecular clock models, as well as different demographic coalescent models of effective bacterial population size (N_e) over time, interpreted as the number of effective infections (i.e. those contributing to onward transmissions)^{30,41}, were tested to infer the demographic history of t008 MRSA strains in northeast Florida. We evaluated two parametric (constant effective population size and exponential population growth), and one nonparametric estimate (Bayesian skyline plot) of bacterial population size over time. The Bayes Factor (BF) strongly favored the relaxed over the strict molecular clock model (BF = 42.4), indicating that different bacterial strains evolved at significantly different rates (Table 2). In addition, analysis

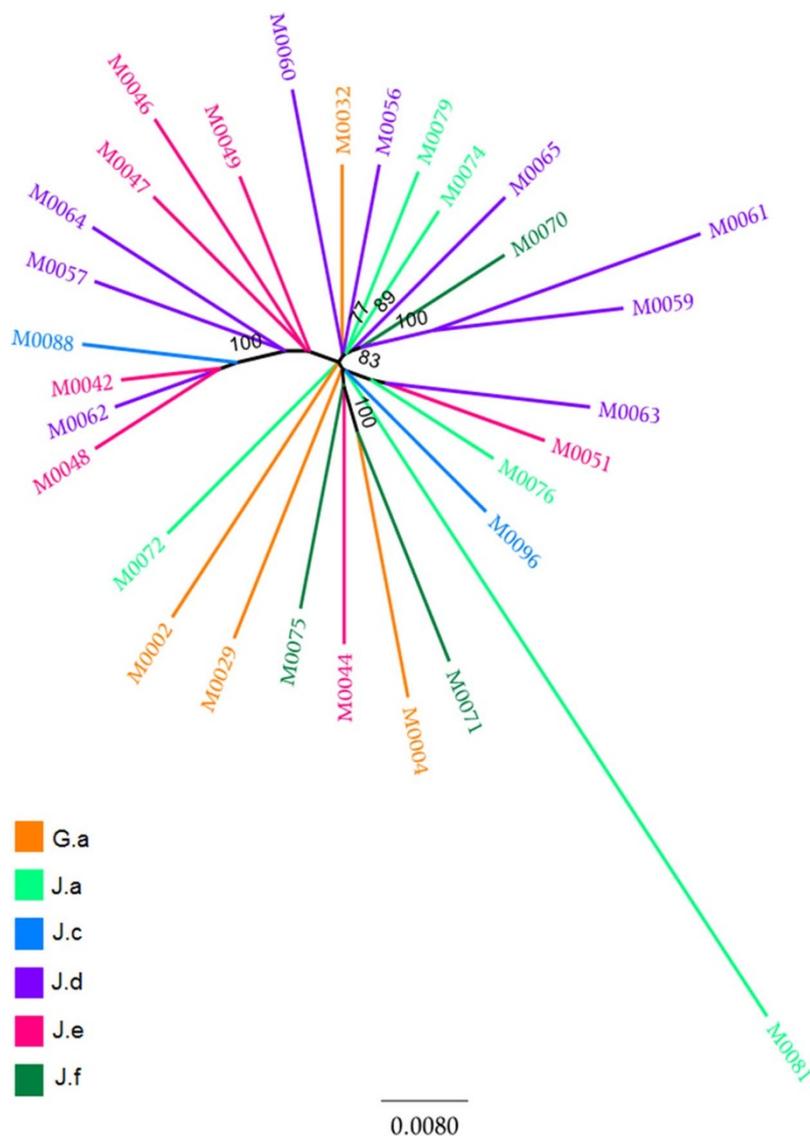


Figure 2 | ML phylogenetic analyses of MRSA t008 in northeast Florida by HCF. Colored tip branches correspond to healthcare facility from which the isolate was obtained. The numbers along the monophyletic branches correspond to bootstrap values (500 replicates). Branch lengths in nucleotide substitutions per site were scaled according to the bar at the bottom of the tree.


Table 2 | Bayes factor between strict (SC) and relaxed (RC) molecular clock

Clock ^a	Marginal likelihood	BF ^b
SC	-9843.445	
RC	-9822.237	42.416

a. The selected molecular clock model (H_0 = Null hypothesis, H_1 = Alternative hypothesis) is highlighted in gray.
 b. BF = Bayes Factor. $6 > BF > 2$ indicates positive evidence against the null hypothesis; $10 > BF > 6$ indicates strong evidence against the null hypothesis, $BF > 10$ indicates very strong evidence against the null model.

of the three demographic models showed positive evidence against the null hypothesis of constant bacterial population size in favor of the exponential growth model (BF = 2.7), which also outperformed the Bayesian skyline plot model (BF = 3.9) (Table 3). The temporal scale of MRSA evolution was inferred using an independent estimate of MRSA genome-wide SNPs evolutionary rate of 7.57×10^{-5} , with a 95% highest posterior density (95% HPD) interval of $5.11 - 10.2 \times 10^{-5}$ nucleotide substitution per SNP site per year³². The reconstruction of MRSA t008 demographic history estimated the origin of the epidemic in mid-1960s, followed by an exponential increase in effective population size, consistent with the known epidemiology of MRSA in the United States (Figure 3). By using the estimated growth rate of MRSA from the exponential population growth model (0.34, 95% HPD = 0.14 - 0.57) and estimates of colonization and infection duration ranging from 8.5 months to one year, we were able to determine the potential reproductive number (R_0) of *S. aureus* within our sample^{42,43}. R_0 estimates were not based on traditional SIR epidemic models; instead, we utilized the estimated N_e and growth rate from phylodynamic analysis to determine the R_0 for infections and colonizations in the population. This method has previously been employed by Pybus et al. to estimate the R_0 of Hepatitis C Virus⁴⁴. R_0 estimates for MRSA USA300 R_0 ranged from 1.24 (95% HPD 1.10 - 1.40) to 1.34 (95% HPD 1.14 - 1.57).

To assess the phylogeographic pattern of MRSA t008 in northeast Florida, a discrete character corresponding to each isolate's respective hospital was assigned to the tip branches of the ML genealogy. Bacterial gene flow among hospitals was then traced on the basis of the maximum parsimony reconstruction of the ancestral characters (Figure 4a). A randomization test showed that the null hypothesis of *panmixia*, i.e. absence of MRSA t008 population subdivision among different hospitals, could not be rejected (Figure 4b). In addition, the observed bacterial gene flow among the different hospitals was not statistically significant (Table 4), indicating a relatively homogenous epidemic across northeast Florida with no directional gene flow between specific hospitals.

Discussion

This study was designed to provide an initial "snapshot" of MRSA distribution and phylogeny within a major metropolitan area, and to permit comparison of these isolates with isolates from a smaller neighboring city. What emerges is the power of coupling high-resolution phylogenetics and phylogeography with genome-wide SNP

data to test molecular epidemiology hypotheses of bacterial spread within a localized healthcare network. These methods can be leveraged to identify emerging epidemics, detect outbreaks, and study antibiotic resistance and virulence, as has been recently demonstrated for *S. aureus*^{45,46}. As rapid sequencing technologies are refined and bioinformatic tools are operationalized, we will continue to observe more examples of the utility of WGS and phylogenetic analysis in routine practice. Ultimately, our findings highlight the clear utility of such methods in seeking to understand, and control, community and regional spread of resistant microorganisms.

We identified a statistically significant difference in the overall distribution of *spa*-types t008 and t002 among a convenience sample of clinical MRSA isolates from Gainesville and Jacksonville HCFs (Figure 1). MRSA *spa* type t008 accounted for 71% of isolates obtained from Jacksonville compared to 16% of strains from Gainesville (Table 1). This is a striking increase from the studies conducted in 2003–2004³, where it was found that t008 strains accounted for only 20% of the isolates, while t002 made up the majority of isolates. Additionally, a national average of 31.3% of t008/USA300 were reported in invasive infections from participating ABC regions reported by the CDC in 2008⁶. These differences must be interpreted with care, however, as they may be attributable to the HCFs' respective patient populations or our sampling strategy. Unexpectedly, phylogeographic analysis of t008 strains demonstrated a lack of clustering, i.e. no hospital-specific clades (Figure 2). It was hypothesized that our analysis would identify monophyletic branches of isolates clustering within hospitals, signifying endemic transmission or distinct community-based sub-epidemics in populations constituting the facility's patient population. The lack of clustering within facilities alludes to other transmission dynamics at work. It was also expected that gene flow of t008 isolates, as measured by the number of observed bacterial migrations in the phylogenetic tree, would be observed between specific HCFs, either in close proximity and/or serving similar populations. However, this was not apparent among our sample, where bacterial strains were randomly distributed among different hospitals without any restricted or directional flow. It is possible that the complexity of the healthcare network within our study area explains the diversity in hospital distribution among phylogenetic clades and the lack of gene flow. For example, the referral system between hospitals and the presence of numerous long-term care, rehabilitation, and long-term acute care facilities is so extensive and geographically dispersed that sub-epidemics propagated through the intermixing of HCF patient populations and the community are spatiotemporally isolated and require a larger sample size to detect. This model would suggest multiple CA and HA reservoirs intermixing and contributing to the overall microbial burden on the HCF. The hypothesis is also supported by the exponential growth of N_e inferred from the MRSA phylogeny utilizing the Bayesian coalescent framework (Figure 3). N_e is a measure of genetic diversity and can be interpreted as the number of bacterial genomes effectively contributing to the subsequent generation⁴⁷. In regards to MRSA, N_e is expected to correlate with the number of infected and/or colonized individuals. However, it is important to note that since individuals may be

Table 3 | Bayes factor between different coalescent models (constant population size, exponential growth, and Bayesian skyline plot)

Demographic model ^a	Marginal likelihood	S.E.	BF ^b		
			Constant vs. Exponential	Constant vs. BSP	Exponential vs. BSP
Constant	-9822.2	+/- 0.25	2.7	1.2	3.9
Exponential	-9820.9	+/- 0.24			
BSP	-9822.8	+/- 0.24			

a. The selected coalescent model is highlighted in gray.

b. BF = Bayes Factor. $6 > BF > 2$ indicates positive evidence against the null hypothesis; $10 > BF > 6$ indicates strong evidence against the null hypothesis, $BF > 10$ indicates very strong evidence against the null model.

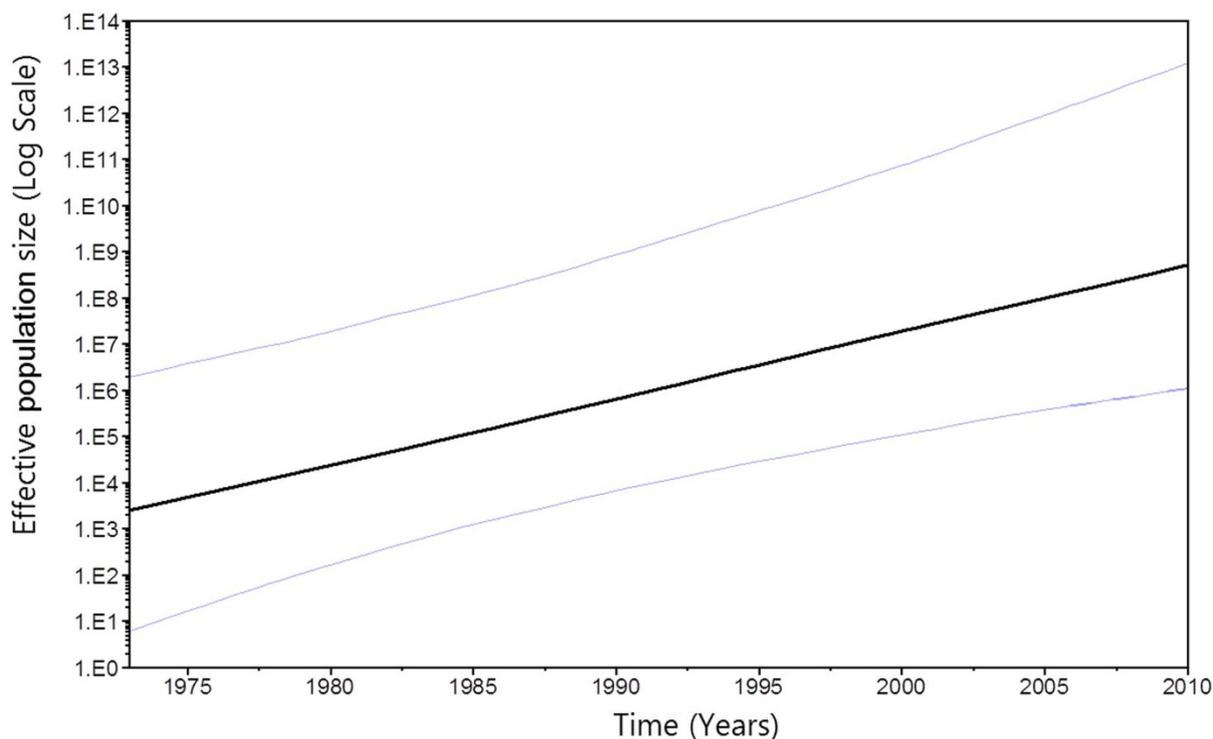


Figure 3 | Bayesian skyline plots of MRSA t008 in Jacksonville. Non-parametric curves of MRSA effective population size (N_e) over time were estimated by employing a Bayesian framework. Genetic distances were transformed into a timescale of years by enforcing a relaxed molecular clock model. Solid lines indicate median (blue), and 95% upper and lower high posterior density (HPD) estimates of N_e (black).

simultaneously infected and/or colonized with multiple strains or not transmit the bacteria, such a correlation is not necessarily 1 : 1³⁰.

By reconstructing the demographic history of bacterial population, it was also possible to estimate R_0 value for MRSA. *S. aureus* demonstrates several clinical presentations ranging from SSTIs to severe invasive disease. Additionally, attempts to accurately estimate R_0 values are confounded by the existence of colonization states, during which individuals can be transiently or persistently colonized by the bacteria. R_0 estimates have ranged from 0.60 to 1.0^{18,42,48}. Our estimates are slightly higher, in agreement with previous evidence that the reproductive rate for CA-MRSA strains may be, in fact, higher than that of HA-MRSA strains¹⁸.

Overall, our study shows the significant implications of utilizing phylodynamic methods for the study of MRSA t008 molecular epidemiology. While we were limited by lack of clinical and demographic data from participating HCFs, our analysis included representative clinical isolates obtained from a similar time point and from all acute care facilities serving a population of well over one million persons within a common geographic region. A recent point prevalence study of HAIs conducted in Jacksonville, FL in 2009 identified 6.0% of patients with one or more HAIs on the day of the survey⁴⁹. *S. aureus* was the most common pathogen, causing 15.5% of HAIs. These results were similar to national prevalence estimates obtained 30 years prior from the Study of Efficacy of Nosocomial Infection Control⁵⁰. Contrastingly, other studies have demonstrated that the incidence of *S. aureus* HAIs in certain geographic areas have been steadily decreasing since 2005⁵¹. These variations in MRSA-related HAI incidence may result from varying infection prevention practices, surveillance methods, or strain distribution. Overall, HAI incidence remains high despite the widespread implementation of comprehensive infection control programs. In this setting, application of this phylodynamic/phylogeographic approach provides a powerful tool to tease apart the various components of MRSA transmission within communities, which is essential for the development of effective interventions.

As in any other study based on epidemiological models, it is important to highlight some of the potential limitations of our approach. First, results should be interpreted in light of our convenience sampling strategy, which increased participation from HCFs at the cost of detailed clinical and epidemiological information. A deliberately planned isolate sampling strategy in future studies will be necessary to strengthen the validity of these initial findings. However, the observation that individuals hospitalized and/or utilizing the emergency departments were infected with a heterogeneous MRSA population remains valid whether the isolates originated from inpatient or outpatient populations. Second, although coalescent-based phylodynamic analysis does not require a large sample size, uniform spatiotemporal sampling is important^{47,52,53}, and it is possible that our convenience sample did not meet this criterion. Finally, the demographic models based on phylodynamic analysis used here assume no population structure and neutral evolution. The violation of either assumption can bias the reconstruction of the population demographic history and ultimately our inference of R_0 . However, phylogenetic and gene flow analysis did show an overall panmictic population (i.e. no population subdivision), suggesting that at least the first of these assumptions may hold true for the present work.

In conclusion, our findings suggest complex transmission dynamics of MRSA between the community and healthcare setting with a heterogeneous distribution of isolates across healthcare facilities. The overall picture of MRSA emergence and distribution drawn herein provides a base for subsequent studies incorporating a structured sampling strategy, phylogenetic analysis of genome-wide SNP data, and corresponding clinical and demographic data. Most importantly, our analysis shows how the application of phylodynamics to microbial genomics has the potential to track the emergence and demographic history of MRSA strains, testing molecular epidemiology hypotheses at a resolution difficult to obtain with alternative molecular typing methods. Microbial phylodynamic analysis may be employed to inform control strategies to prevent nosocomial

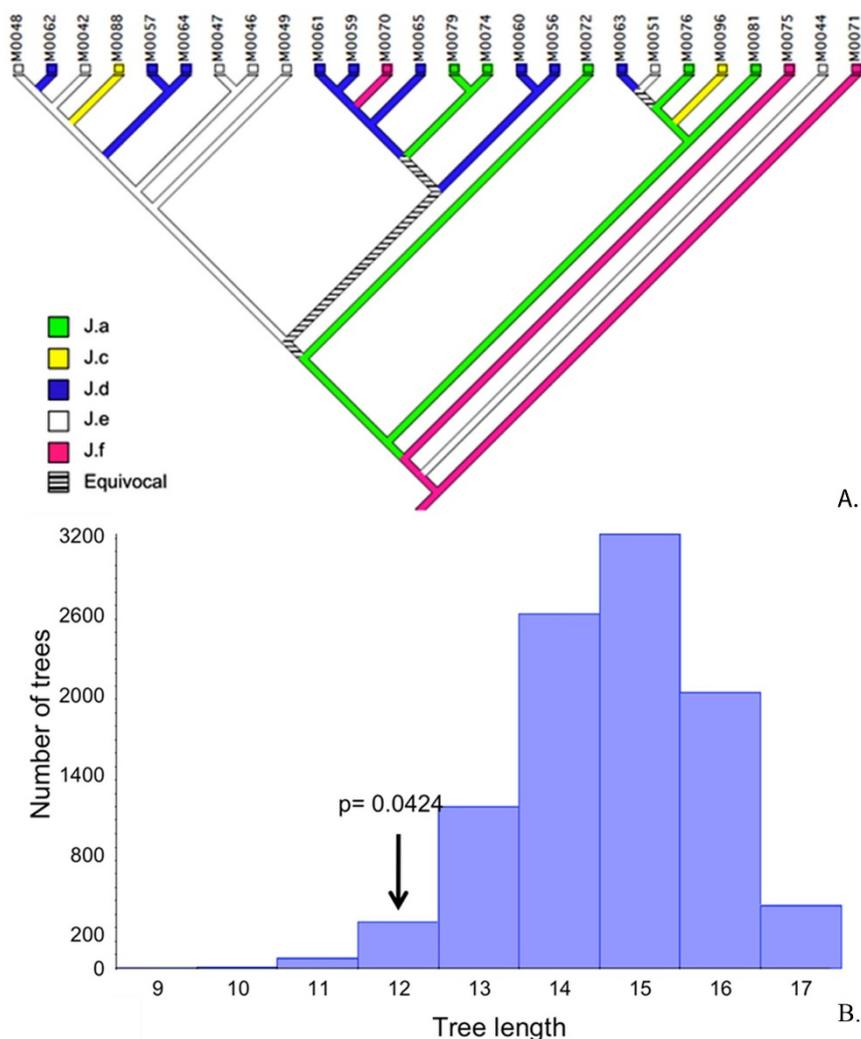


Figure 4 | MRSA t008 phylogeographic patterns in Jacksonville. Phylogeographic analysis using a rooted ML genealogy inferred for 26* sequences from Hospitals J.a, J.c, J.d, J.e, and J.f. A. The most parsimonious reconstruction (MPR) of the state of origin for each internal node (ancestral sequence) in the tree is indicated by the color of the subtending branch according to the legend in the figure. Equivocal branches indicate multiple MPRs. *Note: The four G.a sequences were omitted from the ML genealogy. B. Tree length distribution of 10,000 trees obtained by random joining-splitting. The arrow points to the number of observed migrations in the ML tree.

transmission and/or constant reintroduction of CA-MRSA strains from the community into HCFs.

Methods

Data collection and ethics statement. We obtained a convenience sample of clinical MRSA isolates from invasive and non-invasive infections sent for microbiological culture at six tertiary acute care hospitals in Jacksonville, FL (representing all major hospital systems) and one tertiary acute care hospital in Gainesville, FL during a one week period in September of 2010. Jacksonville is located on the northeast coast of Florida, with a population of approximately 1.3 million in the greater Jacksonville

metropolitan area; Gainesville, with a population of 125,000, is in north-central Florida, approximately 70 miles to the southwest of Jacksonville. Isolates were completely de-identified, with no available clinical or demographic information. The study protocol was reviewed and approved by the University of Florida Institutional Review Board.

Sample processing and spa typing. Cultures of MRSA isolates were processed and an aliquot was frozen at -80°C . Each isolate was grown in liquid subculture overnight at 37°C . Genomic DNA (gDNA) was isolated from pelleted bacteria using the Roche High Pure PCR kit following the standard protocol for isolation of nucleic acids from bacteria (Roche Applied Science, Indianapolis, IN). The quality of the gDNA extracted was determined through gel electrophoresis and quantity was determined using the Nanodrop 2000 (Fisher ND-2000). Molecular typing of these strains was done by *spa* typing. PCR was used to amplify the *spa* repeat region using $1\ \mu\text{l}$ gDNA, 1XGoTaq Green Master Mix and 10pmol of each primer, *spa*-1113f (5'- TAA AGA CGA TCC TTC GGT GAG C -3') and *spa*-1514r (5'- CAG CAG TAG TGC CGT TTG CTT -3'). The primers are numbered from the 3' end of the primer on the forward strand of a reference *S. aureus* sequence (GenBank accession no. J01786; *spa*-1113f [1092-1113] & *spa*-1514r [1534-1514]). Thermal cycling conditions consisted of a hot start (5 min at 80°C) followed by 35 cycles of denaturation (15 s at 94°C), annealing (30 s at 58°C), and extension (60 s at 72°C), with a single final extension of 10 min at 72°C . Gel electrophoresis was used to determine PCR success and all PCR products were Sanger sequenced in both directions. Nucleotide sequences were analyzed by using Ridom StaphType software and synchronized with SpaServer (www.spaserver.ridom.de) in order to assign the *spa* type according to number of tandem repeats and length variation in the *spa* gene⁵⁴.

Table 4 | MRSA metapopulation structure test

Migration direction	Observed gene flow ^a	p-value ^b
J.e-J.d	2	0.9474
J.e-J.c	1	0.9812
J.d-J.f	1	0.8020
J.f-J.e	1	0.9988
J.f-J.a	1	0.9992
J.a-J.c	1	0.9972

a. Number of migrations observed in the ML tree.

b. $p > 0.01$ indicates that the null hypothesis of panmictic spread (not directional gene flow among different hospitals) cannot be rejected.



Next-generation sequencing and data analysis. To provide better discrimination within *spa* type t008 isolates, we conducted next generation WGS. After confirmation of quality and determination of quantity, 5 µg of each isolate gDNA was sequenced on the Illumina HiSeq 2000 sequencing system. Libraries were constructed for each isolate using the Covaris E220 and the SPRIworks Fragment Library System (Beckman Coulter), and accurate quality control was performed using the Agilent Bioanalyzer and qPCR, ensuring library quality before sequencing and ideal cluster density during sequencing. Isolates were uniquely tagged and combined in one of eight lanes of the flow cell for a paired-end 75 base pair read. A *S. aureus* reference strain was identified by querying available WGS in GenBank and utilizing the species tree to select the most appropriate sequence for assembly. After de-multiplexing, single FASTQ output files were mapped to a *S. aureus* reference strain (GenBank accession no. 87125858) with the software BWA version 0.5.8⁵⁵ to obtain SAM files. SAM files were then processed with the Picard and samtools software⁵⁶ into BAM and pileup files, including consensus sequences, using the default parameters for quality calls. Consensus sequences of each strain were merged in a FASTA file, with an additional set of 10 t008 *S. aureus* sequences retrieved from GenBank (accession no. 87159884, 151220212, 57650036, 87125858, 160367075, 87201381, 57284222, 150373012, 88193823, 161508266). Sequences were then aligned with progressiveMauve⁵⁷, and the resulting multiple alignment column positions containing at least a non-ambiguous, non-deletion and non-insertion base change from the consensus (i.e. SNPs) were concatenated in a final alignment.

Phylogenetic and phylogeographic analyses. Phylogenetic signal was assessed by likelihood mapping using TreePuzzle⁵⁸, a transitions/transversions vs. divergence graph, and Xia's test of substitution saturation, as implemented in DAMBE⁵⁹. Recombination evidence was investigated via a phi-test using SplitsTree⁶⁰. Neighbor-joining and ML phylogenies were estimated, assessing node reliability with 500 and 100 bootstrap replicates, respectively, using MEGA5 and PhyML⁶¹. Both the LogDet distance and an optimal model of base substitution (chosen upon the Akaike information criterion) were tested. MRSA t008 demographic history was inferred as previously described by Gray and collaborators³². The hypothesis of metapopulation structure, i.e. the existence of different MRSA sub-populations in distinct hospitals, was tested with a modified version of the Slatkin and Maddison test using the ML tree⁶². The bacterial gene flow (migration) among different hospitals was traced using the state changes and stasis tool (MacClade software), which counts the number of changes in a tree for each pair-wise character state.

Phylogenetic analysis. Different demographic models of MRSA spread were compared using the Bayesian framework implemented in BEAST version 1.7, which employs a Markov Chain Monte Carlo (MCMC) algorithm⁶³. Two parametric (constant population size and exponential growth) and one non-parametric (Bayesian skyline plot) demographic model were compared by enforcing either a strict or a relaxed molecular clock. The evolutionary rate of MRSA genome-wide SNPs estimated by Gray et al. on an independent data set was used as the rate prior³². For each analysis an MCMC was run for 100,000,000 generations with sampling every 10,000th generation. The results were visualized in Tracer v1.4.1. The effective sample size (ESS) value for each parameter was > 300 indicating sufficient mixing of the Markov chain. Model comparison was performed by calculating the Bayes Factor (BF), which is the ratio of the marginal likelihoods (marginal with respect to the prior) of the two models being compared⁶⁴. Approximate marginal likelihoods for each coalescent model via importance sampling using the harmonic mean of the sampled likelihoods (with the posterior as the importance distribution)⁶⁵. Evidence against the null model (i.e. the one with lower marginal likelihood) is assessed in the following way [39]: $2 > 2 \cdot \log_e(\text{BF}) > 6$ indicates positive evidence against the null model; $6 > 2 \cdot \log_e(\text{BF}) > 10$ indicates strong evidence against the null model; $2 \cdot \log_e(\text{BF}) > 10$ indicates very strong evidence against the null model.

R0 estimate. The population growth rate r , which is one of the two free parameters of the exponential population growth model, can be used to infer the epidemiological quantity R_0 . R_0 is the basic reproductive number (infectivity) of a pathogen, i.e. the average number of secondary infections caused by each primary infected individual. In a pathogen population exponentially growing at rate r , where D is the average duration of infectiousness, it can be shown that if the pathogen is transmitted at the same rate during the total length of infection, then $R_0 = rD + 1$ ⁶⁴.

- Diekema, D. J. et al. Survey of infections due to *Staphylococcus* species: frequency of occurrence and antimicrobial susceptibility of isolates collected in the United States, Canada, Latin America, Europe, and the Western Pacific region for the SENTRY Antimicrobial Surveillance. *Clinical Infectious Diseases* **32** Suppl 2, S114–32 (2001).
- Klein, E., Smith, D. L. & Laxminarayan, R. Hospitalizations and deaths caused by methicillin-resistant *Staphylococcus aureus*, United States, 1999–2005. *Emerging Infectious Diseases* **13**, 1840 (2007).
- Jarvis, W. R., Schlosser, J., Chinn, R. Y., Tweeten, S. & Jackson, M. National prevalence of methicillin-resistant *Staphylococcus aureus* in inpatients at US health care facilities, 2006. *American Journal of Infection Control* **35**, 631–7 (2007).
- Klevens, R. M. et al. Changes in the epidemiology of methicillin-resistant *Staphylococcus aureus* in intensive care units in US hospitals, 1992–2003. *Clinical Infectious Diseases* **42**, 389–91 (2006).
- Hubben, G. et al. Modelling the costs and effects of selective and universal hospital admission screening for methicillin-resistant *Staphylococcus aureus*. *PLoS one* **6**, e14783 (2011).
- Centers for Disease Control and Prevention Active Bacterial Core Surveillance (ABCs) Report Emerging Infections Program Network Methicillin-Resistant *Staphylococcus aureus*, 2008. Program (2008) <<http://www.cdc.gov/abc/s/reports-findings/survreports/mrsa08.pdf>> Accessed: November 2012.
- Milstein, A. Ending extra payment for “never events”—stronger incentives for patients’ safety. *The New England Journal of Medicine* **360**, 2388–90 (2009).
- DeLeo, F. R. & Chambers, H. F. Reemergence of antibiotic-resistant *Staphylococcus aureus* in the genomics era. *The Journal of Clinical Investigation* **119**, 2464 (2009).
- Klevens, R. M. et al. Invasive methicillin-resistant *Staphylococcus aureus* infections in the United States. *JAMA: the Journal of the American Medical Association* **298**, 1763–71 (2007).
- Otter, J. A. & French, G. L. Community-associated methicillin-resistant *Staphylococcus aureus* strains as a cause of healthcare-associated infection. *The Journal of Hospital Infection* **79**, 189–93 (2011).
- David, M. Z. & Daum, R. S. Community-associated methicillin-resistant *Staphylococcus aureus*: epidemiology and clinical consequences of an emerging epidemic. *Clinical Microbiology Reviews* **23**, 616–87 (2010).
- Deleo, F. R., Otto, M., Kreiswirth, B. N. & Chambers, H. F. Community-associated methicillin-resistant *Staphylococcus aureus*. *Lancet* **375**, 1557–68 (2010).
- Otto, M. Basis of virulence in community-associated methicillin-resistant *Staphylococcus aureus*. *Annual Review of Microbiology* **64**, 143–62 (2010).
- Li, M. et al. Evolution of virulence in epidemic community-associated methicillin-resistant *Staphylococcus aureus*. *Proceedings of the National Academy of Sciences* **106**, 5883 (2009).
- Cheung, G. Y. C., Wang, R., Khan, B. A., Sturdevant, D. E. & Otto, M. Role of the accessory gene regulator *agr* in community-associated methicillin-resistant *Staphylococcus aureus* pathogenesis. *Infection and Immunity* **79**, 1927–35 (2011).
- Fridkin, S. K. et al. Methicillin-resistant *Staphylococcus aureus* disease in three communities. *New England Journal of Medicine* **352**, 1436–1444 (2005).
- Gonzalez, B. E. et al. Community-associated strains of methicillin-resistant *Staphylococcus aureus* as the cause of healthcare-associated infection. *Infection Control and Hospital Epidemiology: the official journal of the Society of Hospital Epidemiologists of America* **27**, 1051–6 (2006).
- D’Agata, E. M. C., Webb, G. F., Horn, M. A., Moellering, Jr., R. C. & Ruan, S. Modeling the Invasion of Community-Acquired Methicillin-Resistant *Staphylococcus aureus* into Hospitals. *Clinical Infectious Diseases* **48**, 274–284 (2009).
- Merrer, J., Santoli, F., Vecchi, C. & Tran, B. “Colonization pressure” and risk of acquisition of methicillin-resistant *Staphylococcus aureus* in a medical intensive care unit. *Infection control and Hospital Epidemiology* **21**, 718–723 (2000).
- Harmsen, D. et al. Typing of methicillin-resistant *Staphylococcus aureus* in a university hospital setting by using novel software for *spa* repeat determination and database management. *Journal of Clinical Microbiology* **41**, 5442–5448 (2003).
- McDougal, L. K. et al. Pulsed-field gel electrophoresis typing of oxacillin-resistant *Staphylococcus aureus* isolates from the United States: establishing a national database. *Journal of Clinical Microbiology* **41**, 5113–5120 (2003).
- Okuma, K. et al. Dissemination of new methicillin-resistant *Staphylococcus aureus* clones in the community. *Journal of Clinical Microbiology* **40**, 4289–4294 (2002).
- Maltezou, H. C. & Giamarellou, H. Community-acquired methicillin-resistant *Staphylococcus aureus* infections. *International Journal of Antimicrobial Agents* **27**, 87–96 (2006).
- Seybold, U. et al. Emergence of community-associated methicillin-resistant *Staphylococcus aureus* USA300 genotype as a major cause of health care-associated blood stream infections. *Clinical Infectious Diseases* **42**, 647–56 (2006).
- Patel, M. et al. USA300 genotype community-associated methicillin-resistant *Staphylococcus aureus* as a cause of surgical site infections. *Journal of Clinical Microbiology* **45**, 3431–3 (2007).
- Babouee, B., Frei, R., Schultheiss, E., Widmer, A. F. & Goldenberger, D. Comparison of the DiversiLab repetitive element PCR system with *spa* typing and pulsed-field gel electrophoresis for clonal characterization of methicillin-resistant *Staphylococcus aureus*. *Journal of Clinical Microbiology* **49**, 1549–55 (2011).
- Grenfell, B. T. et al. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **303**, 327–32 (2004).
- Gilbert, M. T. P. et al. The emergence of HIV/AIDS in the Americas and beyond. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 18566–70 (2007).
- Gray, R., Tatem, A., Lamers, S. & Hou, W. Spatial phylodynamics of HIV-1 epidemic emergence in east Africa. *AIDS* **23**, 1–14 (2009).
- Hu e, S., Pillay, D., Clewley, J. P. & Pybus, O. G. Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 4425–9 (2005).
- Salemi, M. et al. Different epidemic potentials of the HIV-1B and C subtypes. *Journal of Molecular Evolution* **60**, 598–605 (2005).
- Gray, R. R. et al. Testing spatiotemporal hypothesis of bacterial evolution using methicillin-resistant *Staphylococcus aureus* ST239 genome-wide data within a bayesian framework. *Molecular Biology and Evolution* **28**, 1593–603 (2011).



33. Harris, S. R. *et al.* Evolution of MRSA during hospital transmission and intercontinental spread. *Science* **327**, 469–74 (2010).
34. Köser, C. U. *et al.* Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *The New England Journal of Medicine* **366**, 2267–75 (2012).
35. Mwangi, M. M. *et al.* Tracking the in vivo evolution of multidrug resistance in *Staphylococcus aureus* by whole-genome sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 9451–6 (2007).
36. Nübel, U. *et al.* A timescale for evolution, population expansion, and spatial spread of an emerging clone of methicillin-resistant *Staphylococcus aureus*. *PLoS Pathogens* **6**, e1000855 (2010).
37. Harris, S. R. *et al.* Evolution of MRSA during hospital transmission and intercontinental spread. *Science* **327**, 469–74 (2010).
38. Köser, C. U. *et al.* Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *The New England Journal of Medicine* **366**, 2267–75 (2012).
39. Fitzgerald, J. R., Sturdevant, D. E., Mackie, S. M., Gill, S. R. & Musser, J. M. Evolutionary genomics of *Staphylococcus aureus*: insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 8821–6 (2001).
40. Holden, M. T. G. *et al.* Complete genomes of two clinical *Staphylococcus aureus* strains: evidence for the rapid evolution of virulence and drug resistance. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 9786–91 (2004).
41. Salemi, M., De Oliveira, T., Ciccozzi, M., Rezza, G. & Goodenow, M. M. High-resolution molecular epidemiology and evolutionary history of HIV-1 subtypes in Albania. *PLoS one* **3**, e1390 (2008).
42. Bootsma, M. C. J., Diekmann, O. & Bonten, M. J. M. Controlling methicillin-resistant *Staphylococcus aureus*: Quantifying the effects of interventions. *PNAS* **103**, 5620–5625 (2006).
43. Scanvic, A. *et al.* Duration of colonization by methicillin-resistant *Staphylococcus aureus* after hospital discharge and risk factors for prolonged carriage. *Clinical Infectious Diseases* **32**, 1393–8 (2001).
44. Pybus, O. G. *et al.* The epidemic behavior of the hepatitis C virus. *Science (New York, N.Y.)* **292**, 2323–5 (2001).
45. Price, J. R., Didelot, X., Crook, D. W., Llewelyn, M. J. & Paul, J. Whole genome sequencing in the prevention and control of *Staphylococcus aureus* infection. *The Journal of Hospital Infection* **83**, 14–21 (2012).
46. Harris, S. R. *et al.* Whole-genome sequencing for analysis of an outbreak of methicillin-resistant *Staphylococcus aureus*: a descriptive study. *The Lancet Infectious Diseases* **3099**, 1–7 (2012).
47. Kingman, J. F. C. The coalescent. *Stochastic Processes and their Applications* **13**, 235–248 (1982).
48. Kajita, E., Okano, J. T., Bodine, E. N., Layne, S. P. & Blower, S. Modelling an outbreak of an emerging pathogen. *Nature reviews Microbiology* **5**, 700–9 (2007).
49. Magill, S. S. *et al.* Prevalence of Healthcare-Associated Infections in Acute Care Hospitals in Jacksonville, Florida. *Infection Control and Hospital Epidemiology* (2012).
50. Haley, R. W. *et al.* Nosocomial infections in U.S. hospitals, 1975–1976: estimated frequency by selected characteristics of patients. *The American Journal of Medicine* **70**, 947–59 (1981).
51. Kallen, A., Mu, Y. & Bulens, S. Health Care–Associated Invasive MRSA Infections, 2005–2008. *JAMA: the Journal of the American Medical Association* **304**, 641–648 (2010).
52. Pybus, O. G., Rambaut, A. & Harvey, P. H. An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics* **155**, 1429–37 (2000).
53. Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution* **22**, 1185–92 (2005).
54. Harmsen, D. *et al.* Typing of methicillin-resistant *Staphylococcus aureus* in a university hospital setting by using novel software for spa repeat determination and database management. *Journal of Clinical Microbiology* **41**, 5442–5448 (2003).
55. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* **25**, 1754–60 (2009).
56. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* **25**, 2078–9 (2009).
57. Darling, A. E., Mau, B. & Perna, N. T. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS one* **5**, e11147 (2010).
58. Strimmer, K. & Von Haeseler, A. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proceedings of the National Academy of Sciences* **94**, 6815 (1997).
59. Xia, X., Xie, Z., Salemi, M., Chen, L. & Wang, Y. An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution* **26**, 1–7 (2003).
60. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**, 254–67 (2006).
61. Tamura, K. *et al.* MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* **28**, 2731–2739 (2011).
62. Switzer, W. M. *et al.* Ancient co-speciation of simian foamy viruses and primates. *Nature* **434**, 376–380 (2005).
63. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **1–5** (2012).
64. Kass, R. Bayes factors. *Journal of the American Statistical Association* **90**, 773–795 (1995).
65. Suchard, M. A., Weiss, R. E. & Sinsheimer, J. S. Bayesian selection of continuous-time Markov chain evolutionary models. *Molecular Biology and Evolution* **18**, 1001–13 (2001).

Acknowledgements

This work was supported by the University of Florida Emerging Pathogens Institute seed grant. M.R. is supported in part by the UF CTSI under a grant by the NIH/NCRR Clinical and Translational Science Award UL1 RR029890. We would like to acknowledge the contributions of the North Florida MRSA Collaborative Group and the following individuals to their invaluable contributions to this effort: Diane Halstead, PhD; Yvette McCarter, PhD; Timothy Sellen, MS; Ann Ruby; and Jane Hata, PhD.

Author contributions

M.P.: Next-Generation sequencing data analysis, phylogenetic analysis, interpretation of results, writing of the manuscript; N.V.: phylodynamic analysis, preparation of figures for manuscript; T.A.: preparation of the manuscript, literature search, interpretation of phylogenetic and epidemiological results; M.R.: study design and coordination, sample collection; D.N.: ample preparation and processing, sequencing and laboratory management, data management; K.R.: study design, sample collection; R.L.C.: interpretation of epidemiological data, preparation of the manuscript; J.J.: study design, sample collection and processing, manuscript review and revision; J.M.J.: study design and coordination, interpretation of results, preparation of the manuscript; M.S.: study design, phylogenetic analysis, interpretation of results, preparation of the manuscript. All authors have read and approved the final manuscript.

Additional information

Supplementary information accompanies this paper at <http://www.nature.com/scientificreports>

Competing financial interests: MP and MS are partially supported by the NIH/NCRR CTSI award to the University of Florida UL1 RR02989, and by the NIH-NINDS grant R01 NS063897-01A2. MR is supported in part by the UF CTSI under a grant by the NIH/NCRR Clinical and Translational Science Award UL1 RR029890. This study was also supported by seed funding from the University of Florida Emerging Pathogens Institute. The authors report no competing financial interests.

License: This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>

How to cite this article: Prosperi, M. *et al.* Molecular Epidemiology of Community-Associated Methicillin-resistant *Staphylococcus aureus* in the genomic era: a Cross-Sectional Study. *Sci. Rep.* **3**, 1902; DOI:10.1038/srep01902 (2013).