Supplementary Tables and Figures

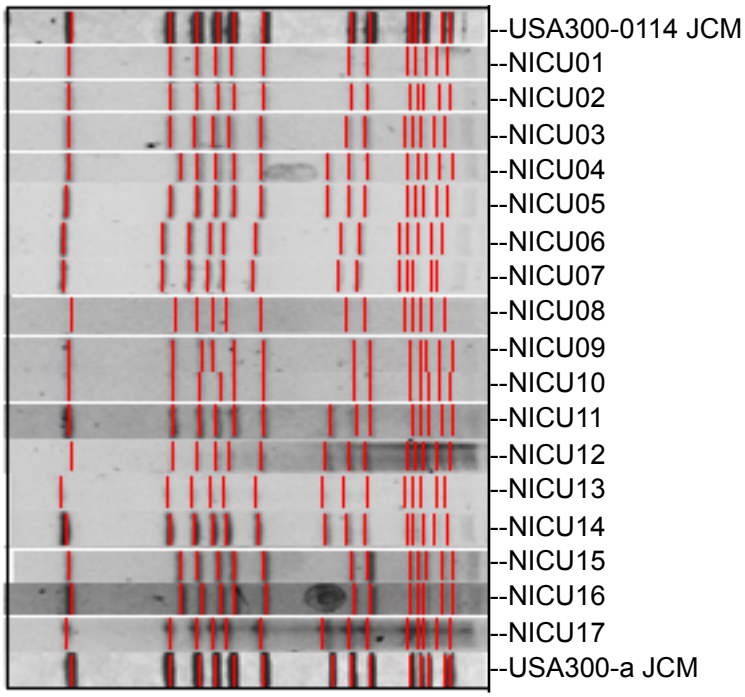
Supplementary Table 1 Antibiotic susceptibilities and *spa*-types of 17 putative outbreak isolates.

Patient	*Spa* Type	Ox	Cc	E	TMP/ SMX	Le	Gn	Rf	Va	Tet	Ln	Dp
1	t008	R	S	R	S	S	S	S	S	S	S	S
2	t008	R	S	R	S	S	S	S	S	S	S	S
3	t008	R	S	R	S	S	S	S	S	S	S	S
4	t5593	R	S	S	S	S	S	S	S	S	S	S
5	t008	R	S	R	S	S	S	S	S	S	S	S
6	t008	R	S	R	S	S	S	S	S	S	S	S
7	t118	R	S	R	S	S	S	S	S	S	S	S
8	t118	R	S	R	S	S	S	S	S	S	S	S
9	t008	R	S	R	S	S	S	S	S	S	S	S
10	t008	R	S	R	S	I	S	S	S	S	S	S
11	t008	R	S	R	S	I	S	S	S	S	S	S
12	t008	R	S	R	S	S	S	S	S	S	S	S
13	t211	R	S	R	S	S	S	S	S	S	S	S
14	t211	R	S	R	S	S	S	S	S	S	S	S
15	t008	R	S	R	S	S	S	S	S	S	S	S
16	t008	R	S	R	S	S	S	S	S	S	S	S
17	t008	R	S	S	S	S	S	S	S	S	S	S

Abbreviations: Ox, oxacillin; Cc, clindamycin; E, erythromycin; TMP/SMX, trimethoprim/sulfamethoxazole; Le, levofloxacin; Gn, gentamicin; Rf, rifampicin; Va, vancomycin; Tet, tetracycline; Ln, Linezolid; Dp, Daptomycin; S, susceptible; I, intermediate; R, resistant

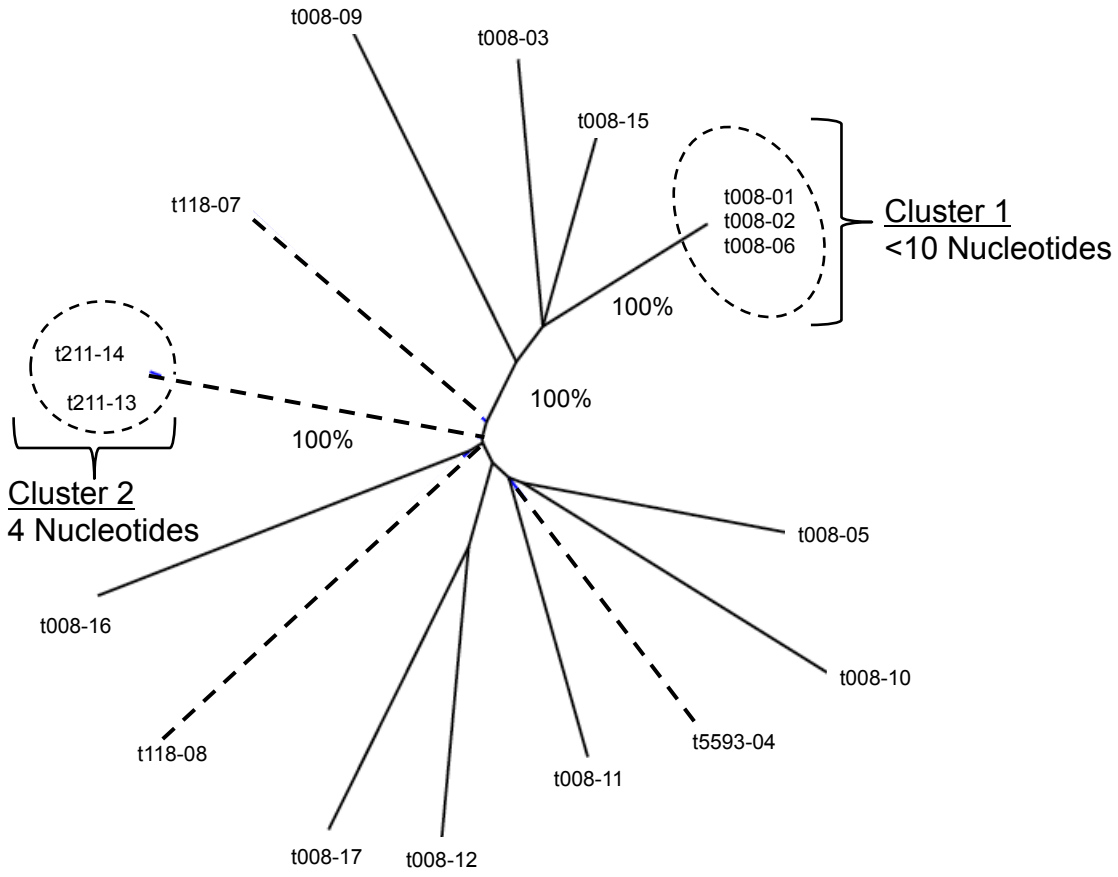
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1																
2	**6**															
3	105	102														
4	133	128	160													
5	137	137	163	129												
6	**10**	**7**	105	123	142											
7	124	122	147	131	143	118										
8	151	147	173	151	158	146	147									
9	129	128	152	170	173	131	159	181								
10	153	148	174	128	135	143	147	165	184							
11	154	150	176	132	141	145	148	165	182	144						
12	152	154	179	153	147	158	158	173	187	167	162					
13	137	133	159	134	145	128	131	150	169	146	144	163				
14	139	137	162	137	147	131	133	154	172	149	148	165	**4**			
15	84	80	106	132	141	77	120	145	130	146	146	158	129	133		
16	146	142	168	144	153	139	140	151	178	158	158	170	143	147	138	
17	150	154	181	153	149	158	160	177	188	167	170	120	165	167	160	172

Supplementary Table 2 Pairwise comparison of nucleotide difference between 17 neonatal intensive care unity methicillin-resistant *Staphylococcus aureus* isolates. There were an average of 143 nucleotide differences (range 4-188) among isolates. We defined nucleotide differences <20 in the core genome as epidemiological linked, which are bolded in the table.



Supplementary Figure 1. Pulse-field gel electrophoresis (PFGE) of Methicillin-resistant *Staphylococcus aureus* (MRSA) isolates from putative outbreak.

PFGE of 17 of 34 MRSA isolates from infants hospitalized in the neonatal intensive care unit belonging to the putative outbreak. PFGE pattern was identified as USA300. References USA300-a and USA300-0114 are included. While PFGE patterns demonstrate variations, dissimilar patterns could not be excluded from consideration of recent transmission.



Supplementary Figure 2. Unrooted maximum likelihood phylogenetic tree illustrating the relationship between 17 putative outbreak isolates. Percentages on branches represent bootstrap support. Tip labels include *spa*-type and patient identification number. Dashed branches depict non-t008 *spa*-types. Longer branch lengths represent more distant phylogenetic relationships and decrease the likelihood of epidemiological linkages during the incident hospitalization. Isolates from patients 1, 2, and 6 differed by less than 10 nucleotides, while isolates 13 and 14, later identified as siblings, different only by four nucleotides.

Supplementary Methods

Whole genome sequencing (WGS), single nucleotide polymorphism (SNP) detection, and phylogenetic analysis

Sample preparation, gDNA isolation, and *spa*-typing was performed as previously described.[1] Isolated gDNA were sequenced using Illumina HiSeq 2000 sequencing system. FASTQ files with raw 2x100 paired-end reads were trimmed and then filtered by Phred quality score of 20 and a minimum length of 30 base pairs using Sickle v. 1.2. Filtered FASTQ files were then mapped to MRSA USA300 reference genome FPR3757 (GenBank accession no. 87125858) with Bowtie 2 v. 2.2.3 short read aligner software using default settings.[2] Local realignment around insertions and deletions was performed using the Genome Analysis Toolkit v. 3.1.1, a framework for analyzing next-generation sequence data.[3] SNPs were called using FreeBayes Bayesian v. 0.9.14 genetic variant detector for haploid organisms using the following settings (--ploidy 1 --left-align-indels --min-base-quality 20 --min-alternate-fraction 0.75).[4] Putative SNPs were initially filtered by depth of coverage (<10) and quality (<20) using VCFtools v. 0.1.10. SNP locations not conserved across all samples (ambiguous sites) were removed. Unmapped reads and mobile genetic elements (e.g. plasmids) were not included in the alignment for phylogenetic analysis as previously described.[5] A FASTA multiple alignment of SNPs was then generated for phylogenetic analysis. The final SNP alignment for the 17 USA300 isolates contained 981 SNPs in the core genome. The bioinformatics pipeline was constructed using the University of Florida High Performance Computing Canter's local instance of Galaxy.[6]

The frequencies of SNP differences were compared between isolates. Recent studies of MRSA intra- and inter-host evolution have suggested that SNP difference ranging from 23-40 are indicative of epidemiological linkages between individuals.[7,8] Considering the relatively short timespan of the epidemic investigation, we conservatively selected a cutoff of 20 SNP differences to indicate a putative transmission event within the NICU. A preliminary neighbor joining (NJ) phylogenetic tree was inferred. The best nucleotide substitution model was then selected using the base tree and a hierarchical likelihood ratio test. Calculations were carried out with MEGA v6.06.[9] A maximum likelihood (ML) tree was then inferred using RAxML v0.7.4, using the GTR+Γ nucleotide substitution model and ascertainment bias correction, with branching patterns evaluated by bootstrapping (1000 replicates).[10]

Bayesian Phylogenetic Analysis

A molecular clock was calibrated to assess the timescale of MRSA spread utilizing the Bayesian framework implemented in the BEAST v1.7.5 package.[11] A Markov Chain Monte Carlo (MCMC) was run for 1 billion generations with sampling every 100,000 generations using the HKY substitution model and the Bayesian skyline plot demographic prior. MRSA evolutionary timescale was estimated by enforcing a lognormal uncorrelated (relaxed) molecular clock where the age for each tip represented by the sampling date of the first positive MRSA isolate from the hospitalized infants. Effective sample size > 200 for each estimated parameter was used as the cutoff to assess proper mixing of the MCMC using Tracer v.1.5. A maximum clade credibility (MCC) tree was selected from the trees posterior distribution using treeAnnotator v1.7.5.

MLST+ Comparison

To explore the feasibility and reproducibility of available software packages with the potential for implementation in the clinical or public health laboratory setting, we analyzed the WGS data using SeqSphere+ software version 2.0 (Ridom, Muenster, Germany). This software conducts a gene-by-gene comparison (MLST+) to produce genome-wide allelic profiles. MLST+ allelic profiles for *S. aureus* were downloaded from the SeqSphere+ server. Assembled genomes in BAM format from our bioinformatics pipeline were imported into SeqSphere+ and allelic profiles were used to create a NJ tree of the 17 NICU isolates. The NJ dendogram from the SNP-based analysis was then compared to the MLST+ NJ dendogram.

References

1 Prosperi M, Veras N, Azarian T, *et al.* Molecular Epidemiology of Community-Associated Methicillin-resistant *Staphylococcus aureus* in the genomic era: a Cross-Sectional Study. *Sci Rep* 2013; **3**: 1902.

2 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25**: 1754–60.

3 McKenna A, Hanna M, Banks E, *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**: 1297–303.

4 Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. 2012; : 9.

5 Harris SR, Feil EJ, Holden MTG, *et al.* Evolution of MRSA during hospital transmission and intercontinental spread. *Science* 2010; **327**: 469–74.

6 Goecks J, Nekrutenko A, Taylor J, Team TG. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* 2010; **11**: R86.

7 Price JR, Golubchik T, Cole K, *et al.* Whole-Genome Sequencing Shows That Patient-to-Patient Transmission Rarely Accounts for Acquisition of *Staphylococcus aureus* in an Intensive Care Unit. *Clin Infect Dis* 2014; **58**: 609–18.

8 Uhlemann A-C, Dordel J, Knox JR, *et al.* Molecular tracing of the emergence, diversification, and transmission of *S. aureus* sequence type 8 in a New York community. *Proc Natl Acad Sci* 2014; : 1401006111 – .

9 Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* 2011; **28**: 2731–9.

10 Stamatakis a. Phylogenetic models of rate heterogeneity: a high performance computing perspective. *Proc 20th IEEE Int Parallel Distrib Process Symp* 2006; : 8 pp.

11 Drummond AJ, Suchard M a, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 2012; : 1–5.